The major histocompatibility complex association in primary sclerosing cholangitis

Sigrid Næss

Norwegian PSC Research Center/Research Institute of Internal Medicine

Division of Cancer Medicine, Surgery and Transplantation

Oslo University Hospital, Rikshospitalet

Faculty of Medicine, University of Oslo

2015

© Sigrid Næss, 2015

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo
No. 2119

ISBN 978-82-8333-142-4

All rights reserved. No part of this publication may be
reproduced or transmitted, in any form or by any means, without permission.

Cover: Hanne Baadsgaard Utigard.
Print production: John Grieg AS, Bergen.

Produced in co-operation with Akademika Publishing.
The thesis is produced by Akademika Publishing merely in connection with the thesis defence. Kindly direct all inquiries regarding the thesis to the copyright holder or the unit which grants the doctorate.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The work presented in this thesis has been conducted at Norwegian PSC Research Center (NoPSC) at Oslo University Hospital, Rikshospitalet from 2010-2015. My deepest gratitude goes to Canica AS who has funded my position and made this research possible.

A warm thank you goes to all patients and healthy controls enrolled in the various studies, without your contribution this research would not have been possible.

A special gratitude goes to my main supervisor, Tom Hemming Karlsen. Thank you for taking a chance on me and giving me the opportunity to work in an excellent scientific environment. Your contribution to PSC research is outstanding and your high standards are inspiring. Thank you for your motivation and support throughout my PhD period, especially through difficult times.

Another special thank you goes to my co-supervisor, Benedicte Alexandra Lie. Your support, supervision and laughter throughout my entire PhD period have been invaluable. You are not only an excellent researcher, but also a genuinely warm person who always find the time to motivate, supervise and support your PhD students no matter how busy you are.

Johannes E. R. Hov (co-supervisor), your everyday availability and guidance in everything from statistics to paper writing have been invaluable to me. Your knowledge, pedagogical skills and humor are highly appreciated.

Espen Melum must be mentioned in particular. Thank you for teaching and introducing me to the world of programming and custom scripts, a world I never thought I would enter. And for your general statistical input and genuine enthusiasm.

Trine Folseraas, for taking the time to proof-read this thesis and for answering questions throughout my PhD-period. You are a great colleague.

Erik Schrumpf and Kirsten Muri Boberg for your contribution to PSC research through so many years. Your inputs to my projects have been highly appreciated.

Each and one of our collaborators and co-authors are acknowledged, your contributions are highly appreciated. A special thank you goes to collaborators at Cambridge Institute for Medical Research; Professor John Trowsdale, Dr. James Traherne, Dr. Wei Jiang and Jyothi Jayaraman, for your guidance and warm and welcoming atmosphere during my visits to

Cambridge. A special thank you to Bjarte Fosby for your collaboration, and for taking time to answer “last minute questions” in a very hectic period of time.

I want to thank the Research Institute of Internal Medicine, Department of Immunology and Department of Medical Genetics, who all have provided me with stimulating work environment during my PhD-period.

Former and present members of the NoPSC group need to be mentioned in particular; Hege, Bente, Jarl Andreas, Mona, Liv Wenche, Kristian, Jorunn, Kim, Bjarte, Espen, Johannes, Trine, Kristin, Alexey, Anders, Elisabeth, Martin, Nathalie, Eva Kristine, Mette, Erik and Kirsten. You all contribute(d) to great research and a great work environment.

I would also like to thank former and present members of the Immunogenetics group including Marte V., Marthe M., Marthe G., Siri, Angelina, Inger-Lise, Gry, Hanne, Hanna, Ingvild, Fatemeh, Kari and Line. Being part of the Immunogentics group has been very important to me, both scientifically and socially.

To Tonje, Maria and Elisabeth; my fellow PhD students and friends. Your presence, humor and friendship have made my PhD-period so much better, both through ups and downs.

To all my lovely friends and family, too many to mention by name, your friendship and love is what really matters in life. A special thank you to mum and dad who have always supported and cheered for me.

To Trygve, my personal IT-support, best friend and husband. You deserve an entire page of acknowledgements, but I will try to keep it short. You are my most important support, without you this would not have been possible. My love for you is endless.

Sigrid Næss

Oslo, 2015

ABBREVIATIONS

ACPA	anti-citrullinated peptide antibodies
AH7.1	ancestral haplotype 7.1
AH8.1	ancestral haplotype 8.1
AIC	akaike information criterion
AIH	autoimmune hepatitis
ANCA	anti-neutrophil cytoplasmatic antibody
CI	confidence interval
CNV	copy number variation
GWAS	genome-wide association study/ies
HLA	human leukocyte antigen
HWE	hardy-weinberg equilibrium
IBD	inflammatory bowel disease
KIR	killer immunoglobulin-like receptor
LD	linkage disequilibrium
MAF	minor allele frequency
Mb	mega or million base pairs
MHC	major histocompatibility complex
MICA	major histocompatibility complex class I chain-related A
NK cell	natural killer cell
OR	odds ratio
PBC	primary biliary cirrhosis
PSC	primary sclerosing cholangitis
QC	quality control
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
UDCA	ursodeoxycholic acid
WGA	whole-genome amplified

LIST OF PUBLICATIONS

Paper I

Naess S, Lie BA, Melum E, Olsson M, Hov JR, Croucher PJP, Hampe J, Thorsby E, Bergquist A, Traherne JA, Schrumpf E, Boberg KM, Schreiber S, Franke A, Karlsen TH. Refinement of the MHC Risk Map in a Scandinavian Primary Sclerosing Cholangitis Population. PLoS ONE, 2014 Dec; 9(12): e114486.

Paper II

Næss S, Björnsson E, Anmarkrud JA, Al Mamari S, Juran BD, Lazaridis KN, Chapman R, Bergquist A, Melum E, Marsh SGE, Schrumpf E, Lie BA, Boberg KM, Karlsen TH, Hov JR. Small duct primary sclerosing cholangitis without inflammatory bowel disease is genetically different from large duct disease. Liver International, 2014 Nov; 34(10):1488-95.

Paper III

Fosby B, Næss S, Hov JR, Traherne J, Boberg KM, Trowsdale J, Foss A, Line PD, Franke A, Melum E, Scott H, Karlsen TH. HLA variants related to primary sclerosing cholangitis influence rejection after liver transplantation. World Journal of Gastroenterology, 2014 Apr; 20(14): 3986-4000.

1 INTRODUCTION

The main topic of this thesis has been the genetic association of the major histocompatibility complex (MHC), also called the human leukocyte antigen (HLA) complex in humans, at chromosome 6p21 in primary sclerosing cholangitis (PSC), with major focus on the classical HLA genes. The genetic predisposition to PSC has been a major priority of PSC research the last decade as a first step towards understanding the molecular basis of the disease. Large scale genome-wide association studies (GWAS) have been performed and identified 16 regions in the genome that are associated with PSC, including the MHC association, which was first detected in the early 1980s by researchers in Oslo and Oxford (1, 2). Following this first discovery, research of the MHC association in PSC has been a priority at Oslo University Hospital, including several doctoral theses. However, the complexity of this genetic region is substantial and has made it difficult to identify true causative genetic variants and thus move forward into molecular translation of this genetic finding. Recent advancement in genotyping technology and statistical tools has made it possible to further characterize genetic associations with the MHC, which was a major aim of this thesis.

First, an introduction to PSC and genetics (with main focus on the MHC) is given. Second, major aims of this thesis are presented. Third, main results from the three papers are given. Fourth, methodological aspects are discussed followed by a general discussion of the results and major challenges related to analysis and interpretation of associations in the MHC. Lastly, conclusions are made and future studies of PSC within this field are outlined.

1.1 Primary sclerosing cholangitis

1.2 Epidemiology

PSC is a progressive, chronic inflammation disorder, affecting both intra -and extrahepatic bile ducts (Figure 1 illustrates a PSC affected liver). Sustained inflammation results in scarring of the bile ducts, blockage of bile flow (cholestasis) and ultimately end-stage liver disease requiring liver transplantation. PSC is a rare condition, with a geographical variation in incidence and prevalence (reviewed in (3) and (4)), with highest frequency in patients of Northern European descent with an incidence of up to 1.3 per 100 000 inhabitants per year (5). Although PSC can be observed in both sexes and at any age, the typical PSC patient is male (a 2:1 male to female ratio) and in early adulthood, with a peak in incidence in the 3th and 4th decade of life. Co-morbidities are common, most frequent is inflammatory bowel disease (IBD) which is observed in ~ 40-80 % of PSC patients, dependent on the ethnicity of the population studied (6-10). Other immune mediated diseases such as type 1 diabetes, autoimmune thyroid disorders, psoriasis and rheumatoid arthritis are observed in up to 25 % of the patients (11). PSC patients are at higher risk of developing hepatobiliary malignancies (i.e. cholangiocarcinoma, hepatocellular carcinoma and gallbladder carcinoma), in particular cholangiocarcinoma, with a lifetime risk of ~ 10-15 % (12-14). Increased risk of colorectal cancer (in patients with IBD) has also been reported (15, 16). Although a rare disease, PSC is important as it affects young people and reduces both quality of life and life expectancy.

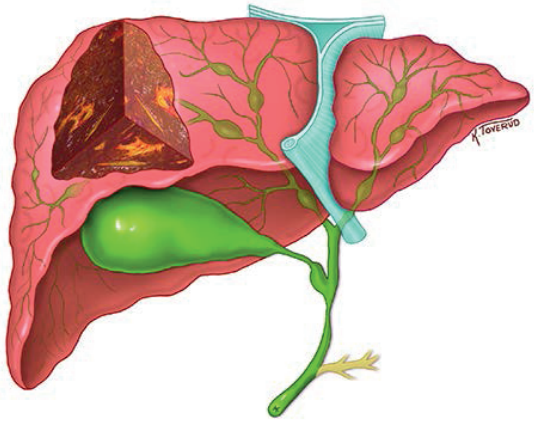

Figure 1: Illustration of a liver and its biliary tree affected by primary sclerosing cholangitis.
Several strictures with intervening saccular dilatations of both the intra –and extrahepatic bile ducts are commonly observed. The strictures lead to cholestasis, peribiliary fibrosis and eventually liver cirrhosis. The figure was created by Kari C. Toverud, CMI. (Copyright K. Toverud).

1.2.1 Clinical presentation, diagnosis, progression and treatment

The clinical presentations of PSC vary. Most common signs and symptoms upon diagnosis are fatigue, pruritus, jaundice, hepatomegaly and abdominal pain (Table 1) (reviewed in (17)). Still, a large proportion (as many as 57 %) of the patients are reported to be asymptomatic at time of diagnosis (18).

Table 1: Signs and symptoms of primary sclerosing cholangitis at diagnosis, adapted from (17).

Signs and symptoms	Prevalence
Fatigue	43-75 %
Pruritus	25-59 %
Jaundice	30-69 %
Hepatomegaly	34-62 %
Abdominal pain	16-37 %
Splenomegaly	14-30 %
Hyperpigmentation	25 %
Weight loss	10-34 %
Variceal bleeding	2-14 %
Ascites	2-10 %

A diagnosis of PSC is made in patients with i) elevated serum markers of cholestasis (alkaline phosphatase and γ-glutamyl transferase), not otherwise explained, ii) classical cholangiographic findings and iii) no evidence of secondary cause of sclerosing cholangitis (19). Elevations of serum markers of cholestasis are often observed at diagnosis, in particular alkaline phosphatase, which is observed in 84-92 % of cases (7, 20). As with signs and symptoms, cholestatic markers can also be within the normal range, this is often due to the fluctuating nature of these parameters during the course of disease. Imaging findings observed with magnetic resonance cholangiography or endoscopic retrograde cholangiography are the gold standard for the diagnosis of PSC. Affection of large intrahepatic bile ducts and/or extrahepatic bile ducts is observed, and a typical "beaded" appearance due to multifocal strictures with proximal dilatations is visualized (Figure 2). Secondary causes of sclerosing cholangitis, such as chronic biliary infections, cholangiocarcinoma, biliary calculi, surgical trauma and IgG4-associated sclerosing cholangitis must be excluded prior to making a diagnosis of PSC. Liver biopsy is not necessary for the diagnosis of PSC in patients with typical cholangiographic findings. Yet it might be valuable to exclude other diagnosis, to support the diagnosis of PSC, and to diagnose the subgroup of PSC patients that only have affection of small intrahepatic bile ducts not visualized by cholangiography (i.e. small duct PSC).

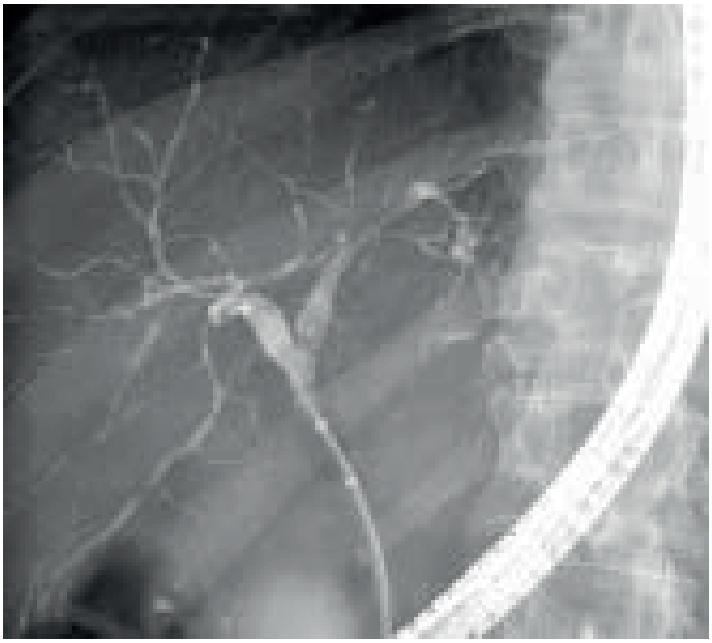

Figure 2: Endoscopic retrograde cholangiography of a primary sclerosing cholangitis patient.
A typical "beaded" appearance with multiple strictures and associated dilatations of the biliary tree is visualized. This figure was kindly shared by Vemund Paulsen.

Various prognostic models for PSC exist and have been useful in predicting outcome in patient cohorts. However, as for the clinical presentation described above, the progression of the disease is highly variable between patients, and it is not recommended to use prognostic models to predict individual outcomes (21). Nevertheless, for most patients the disease is progressive by nature and time from diagnosis to end stage liver disease, represented by time to death or liver transplantation, ranges from 10-18 years (7, 20, 22).

1.2.2 Liver transplantation and medical therapy

No curative or disease halting medical therapy exists for PSC as of to date, and liver transplantation is the only effective treatment currently available for end stage liver disease due to PSC. Indications for liver transplantation in PSC patients are similar to those with other forms of chronic liver disease, including complications related to portal hypertension, impaired quality of life and chronic liver failure. Other indications for liver transplantation in PSC patients include intractable pruritus, recurrent bacterial cholangitis and cholangiocarcinoma. Liver disease risk scores (i.e. UKELD and MELD score) are employed

to identify suitable patients for liver transplantation. Although liver-transplanted PSC patients have a 5-year survival of 75-85 % (23, 24), risk of recurrent disease is present and estimated to occur in ~ 20 % (reviewed in (25)). Medical therapy is for the time being directed against symptom relief and complications. The most intensively studied drug is ursodeoxycholic acid (UDCA). Administrating low to high dosages (10-30 mg/kg/day) of UDCA has led to improvements in serum liver tests, including alkaline phosphatase, alanine and aspartate aminotransferase, bilirubin, albumin and γ-glutamyl transferase, however no impact on survival has been found (26-29). More importantly, high dosages (28-30 mg/kg/day) correlated with adverse events such as liver transplantation and death in one study (30). Based on these observations there are currently controversies whether administration of low dosages of the drug is of benefit or not. The American Association for the Study of Liver Diseases recommends against the use of UDCA, whereas the European Association for the Study of Liver has a more liberal opinion, stating that there is insufficient evidence to make a clear recommendation for time being (19, 21). Novel agents directed at fibrogenesis as well as antibiotic therapy are possible future therapies.

1.2.3 Small duct PSC

Some patients have elevated cholestatic serum markers and liver histology compatible with PSC, but with a normal cholangiogram. These patients are diagnosed with small duct PSC, as opposed to "normal", "classic" or "large duct" PSC in which also large intrahepatic and/or extrahepatic bile ducts, as visualized by cholangiography, are affected. These patients make up about 10% of the overall PSC population (18). The general prognosis of small duct patients is better than for large duct disease, with fewer liver related deaths, longer transplantation-free survival (31-34), and hardly ever development of cholangiocarcinoma (35), without a foregoing progression to large duct disease. Due to this phenotypically milder presentation, it has been debated whether small duct PSC is a milder variant of large duct PSC, or even a distinct disease entity. However, follow-up studies have indicated that ~ 25 % of the small duct population progress to large duct disease with time (34), suggesting that at least a proportion of the patients initially diagnosed with small duct PSC, represent an early stage of large duct PSC. Due to its phenotypical distinction from large duct PSC these patients have been excluded from previous genetic studies of PSC. Because of this, and the rarity of the condition, previous studies of small duct PSC are scarce and limited to follow-up studies focusing on the natural progression of the disease. To better characterize the relationship between large duct and small duct PSC, we conducted the first genetic study in small duct

PSC, focusing on the strongest genetic association found in large duct PSC, i.e. variations within the MHC (paper II).

1.2.4 Etiopathogenesis

The etiology and pathogenesis of PSC are largely unknown. The etiology is most likely multifactorial, meaning that both genetic and environmental factors, and the interplay between these (i.e. gene-gene, gene-environment and environment-environment), play a part. Few environmental factors are known to influence the risk of PSC, except a possible protective effect of smoking and coffee consumption (36-39). Evidence of a genetic contribution to PSC risk (introduced in more detail in section 1.2.5) is now well established through several large scale genetic studies, and involves genes related to immunological pathways, cholangiocyte biology, liver fibrosis and risk of cancer (40-45).

As the initiating factor triggering disease, e.g. a virus or bacteria, may be long gone at time of diagnosis, focus on pathogenic mechanisms has been a priority. There is no universally agreement on pathogenic mechanism(s) responsible for the bile duct injury observed in PSC, and hypotheses involve both immune mediated injury and aspects related to bile physiology. The gut-liver axis and the close relationship between PSC and IBD form the basis of one of the proposed pathogenic mechanisms of PSC, incorporating both "the leaky gut" hypothesis and the "aberrant homing" hypothesis. A "leaky gut", due to colitis, may cause translocation of harmful compounds from the microbiota, e.g. infectious and toxic agents, to the portal system with subsequent activation of the immune system and inflammation of the biliary tree. In addition, it has been shown that mucosal T cells, activated in the setting of colitis, can be misdirected to the liver (i.e. "aberrant" homing) and thus contribute to the inflammatory process observed in PSC (46). A predominant MHC association, presence of autoantibodies, and co-existence of other autoimmune diseases in some patients, have led to the "autoimmune" hypothesis in PSC pathogenesis. However, as PSC has many atypical autoimmune features, including a male predominance, non-specific autoantibodies and non-responsiveness to immunosuppressive therapy, it is debated whether PSC should be classified as an autoimmune disease or an immune mediated inflammatory disease (47). Alternations of bile hemostasis, including abnormal (i.e. "toxic") composition of bile, as well as a defective protection system against bile in the biliary epithelium, are other proposed pathogenic mechanism of PSC. This is exemplified by the cholestatic phenotype observed in humans and mice with defects in the *ABCB4* and *Abcb4* genes, respectively (48, 49), encoding phospholipid transporters (i.e. MDR3 and Mdr2) important for normal bile secretion.

None of the proposed hypotheses rule out one another, and given the diverse nature of PSC, there are most likely several components to PSC etiopathogenesis. However, it is reasonable to argue that PSC occurs due to some sort of environmental trigger(s) in genetic susceptible individuals.

1.2.5 Heritability of PSC

With a 9-39 fold increased risk of disease in siblings of PSC patients (50), there is evidence of a genetic component to this condition. However, the heritability does not show a simple Mendelian inheritance pattern, and PSC is characterized as a complex genetic trait. Complex traits are believed to result from variation within multiple genetic locations in the human genome and their interplay with environmental factors. GWAS have during the past decade facilitated discovery of such genetic variants in numerous of traits, also in PSC. At the beginning of my PhD-period, association with variations within three genetic locations was established for PSC at a genome wide significance level (P-value $< 5 \times 10^{-8}$) (41). As of date, this number has increased to 16 (40, 41, 43, 45). Half of these overlapped with genetic associations observed in IBD (Crohn's disease and ulcerative colitis). Of notice, many of the 16 loci also overlapped with associations of other immune mediated diseases like type 1 diabetes (8/16), coeliac disease (7/16) and rheumatoid arthritis (6/16) (51). The by far strongest genetic association in PSC, first shown in 1982 (1, 2), and replicated numerous times, is located at chromosome 6p21 within the MHC. This genetic region, and its relevance in PSC, form the basis of this PhD thesis and are introduced in greater detail in the following section (1.3).

1.3 Genetics and the Major Histocompatibility Complex

1.3.1 Introduction to genetic terms and genetic variation

Our genetic information (also called the human genome) is encoded by DNA sequences consisting of ~ 3 billion base (i.e. nucleotide) pairs distributed across 23 pairs of chromosomes in the cell nucleus (52, 53). The human genome includes both protein coding regions (i.e. genes) and non-coding regions. Human genetic diversity account for 0.1 % of the genome when any two humans are compared (54). Although quite small in number, it is this variation which is of interest and cause (together with environmental factors) different individual observable traits, also called phenotypes. Examples of phenotypes are; height, eye color, or presence of a disease.

Variations in the human genome range from gross alternations, e.g. number of chromosomes, to smaller nucleotide sequence differences. In the broadest sense, heritable nucleotide sequence differences between individuals can be divided into different groups based on number of nucleotides affected; smaller sequence variants, including single nucleotide variants (SNVs) and short insertions and deletions, and larger sequence variants, referred to as structural variants (55). SNVs are DNA sequence variation in which a single nucleotide is altered between two homologous chromosomes due to a foregoing mutation, it is referred to as a single nucleotide polymorphism (SNP) if a frequency of 1 % or higher is observed in the population (Figure 3). Structural variants are typically classified as sequence variants of at least 50 base pair in size, and include larger insertions and deletions, duplications, translocation and inversions (Figure 4).

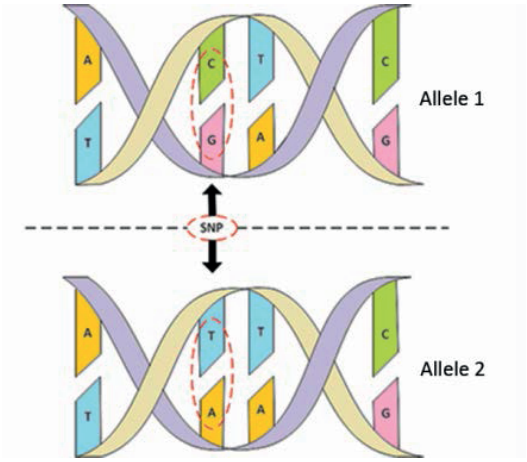

Figure 3: Illustration of a single nucleotide polymorphism.
The two homologous chromosomes differ at one single base pair position (indicated by the red circle), this represents a so-called single nucleotide polymorphism.

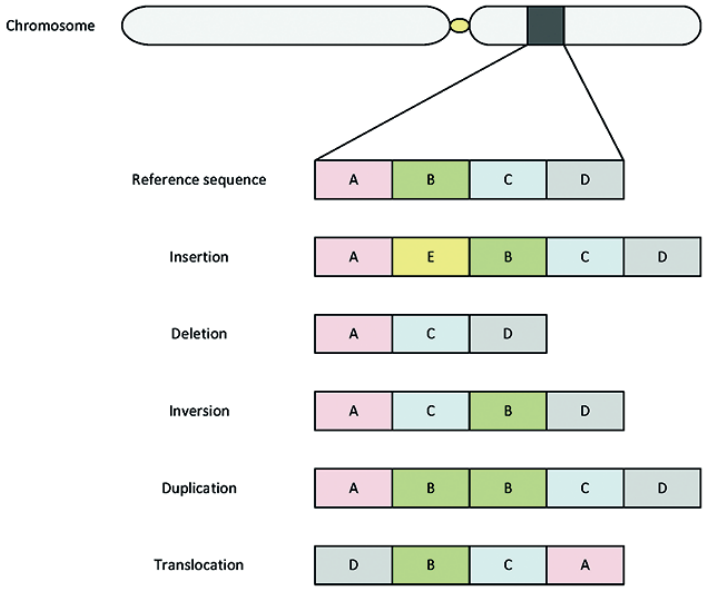

Figure 4: Human genetic structural variations.
Various classes of structural variations are illustrated relative to a reference sequence, including insertion, deletion, inversion, duplication and translocation.

In genetics, a locus (plural loci), is a specific location in the genome. It could either be the location of a gene, which typically stretches over several thousand base pairs, or it could describe the position of a SNP, which point to a single base pair position. An allele is one of a number of alternative forms of the same genetic locus, typically used to describe variant forms of a gene or a SNP. For instance one of the most variable (polymorphic) genes in humans, *HLA-B*, has 3887 different alleles assigned (as of April 2015) (http://www.ebi.ac.uk/ipd/imgt/hla/). In comparison most SNPs are biallelic (i.e. two variant forms). Any individual has two alleles (one on each parental chromosome) at a locus, and this combination of alleles is referred to as a genotype. A genotype can either be homozygous (i.e. the two alleles are identical) or heterozygous (i.e. the two alleles are different).

The distinct combination of alleles at two or more loci present on the same chromosome constitutes a haplotype. Sometimes certain combinations of alleles (i.e. haplotypes) are seen more often or less often than expected based on the observed individual allele frequencies in the population. This non-random association of alleles at two or more loci is called linkage

disequilibrium (LD). Conversely, when two alleles are independent of each other, they are said to be in linkage equilibrium. LD is usually measured by D' and r^2, with absolute values between 0-1. The closer the value is to 1, the stronger the LD. Several factors influence the degree of LD between two loci, but as a general rule the degree of LD in a population will decrease as the population gets older, due to the increasing number of recombination events. Hence, the degree of LD in an African population will be lower than in a population of newer origin, as for instance the Norwegian population. The concept of LD is important in genetics, and is, as will be introduced and discussed later, both an advantage and disadvantage when performing genetic association studies.

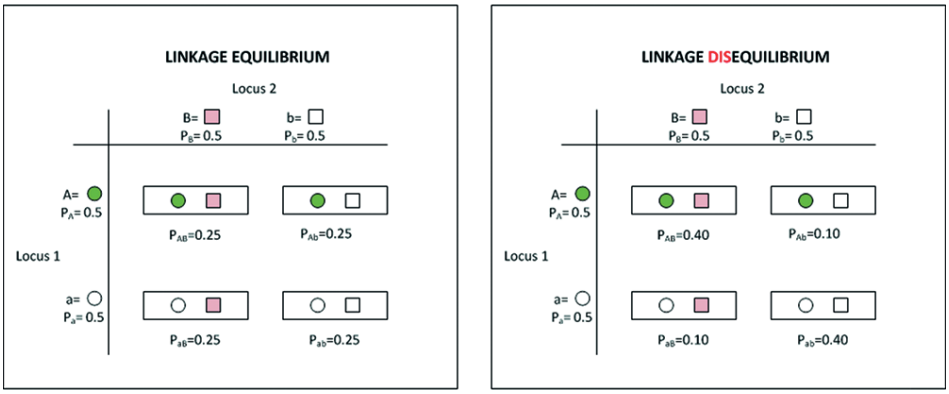

Figure 5: The concept of linkage equilibrium versus linkage disequilibrium.
Observed population frequencies of allele *A* and *a* at locus 1, and allele *B* and *b* at locus 2, are all 50%. If alleles at locus 1 and 2 are completely independent, frequency of the four possible combinations (i.e. haplotypes); *AB*, *Ab*, *aB* and *ab*, is simply given by; P = 0.5 x 0.5. Each of the four haplotypes will then have a frequency of 25%. If the frequency of each allele within the haplotypes is the same as the frequency of that allele in the population as a whole, the alleles are said to be in *linkage equilibrium* (illustrated in the left figure box). The right figure box illustrates the opposite concept, *linkage disequilibrium*, in which the various haplotypes are observed more often or less often than expected by the individual allele frequencies. Figure inspired by graphics at https://www.withfriendship.com/user/neeha/Linkage_disequilibrium.php

1.3.2 From discovery of DNA to whole genome sequencing

In the 1940s DNA was discovered and its three dimensional structure was deduced by Watson and Crick in 1953 (56). Biological insight and technological advances in the 1970s and 80s made it possible to copy DNA and read the order of bases (i.e. DNA sequencing techniques such as "Sanger-sequencing" was established). In the 1990s sequencing of genomes began

and in February 2001 two reference versions of the human genome were published simultaneously by two competing groups (52, 53). However, these reference versions did not annotate genetic variations and so the next goal of the genetic society was to catalogue genetic variations and similarities between humans, in order to explore how genetic variants contributed to phenotypic diversity. Extensive efforts have been made through large-scale international projects such as "the HapMap project" (57, 58) and "the 1000 Genomes project" (59) the last decade to achieve this goal. In addition to characterize genetic variation between humans, both the HapMap data and the 1000 Genomes data have been a tremendous resource for numerous large scale genetic association studies of various complex genetic traits. Based on the "common disease, common variant" hypothesis, which states that common variants with relatively small effect sizes causes common phenotypic traits, hundreds of GWAS have been conducted the last decade. These studies genotype and compare a relatively large number (typically 500.000 to 1.000.000) of common variants (i.e. SNPs) throughout the genome between affected individuals (i.e. cases) and unaffected controls. An even larger number of genetic variants can be inferred through imputation algorithms and the use of public available reference data from the HapMap and 1000 Genomes projects. That is, due to the phenomenon of LD (described in 1.3.1), genotypes of genetic variants not directly genotyped in your study population can be inferred (i.e. imputed) based on genotypes available in a reference data set (see section 4.5 and Figure 11) and thus boosting the number of SNPs available for analysis. Reference data from the 1000 Genomes project was used in paper I to impute SNP genotypes.

1.3.3 The MHC – an important and complex genetic region

The MHC is located on the short arm of chromosome 6, 6p21.3, and span ~4 Mb (million base pairs) incorporating ~260 genes (60).

The MHC is divided into three subgroups; class I, class II and class III (Figure 6). About 30 % of the genes located within the MHC encode proteins of immunological function, including antigen processing and presentation, cytokines and chemokines. The most studied, are the so-called classical HLA genes. These are located within the MHC class I and class II region, and referred to as HLA class I and II genes. The current nomenclature of these genes and their alleles is presented in Figure 7.

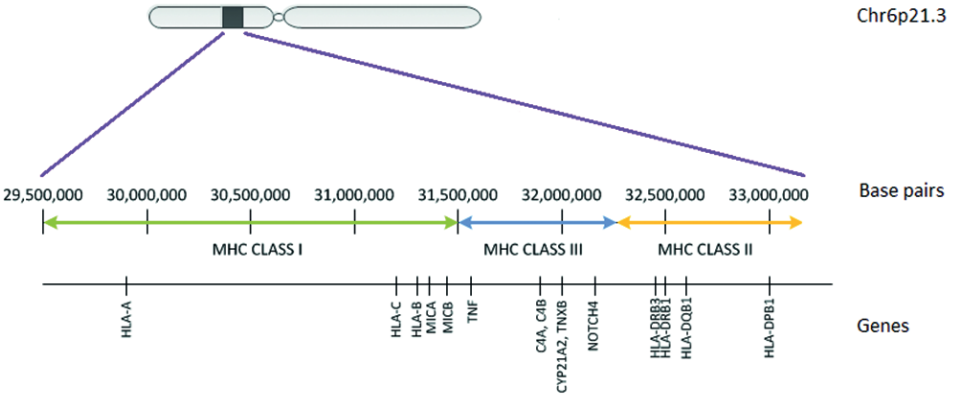

Figure 6: Major histocompatibility complex organization.

The classical major histocompatibility complex (MHC) is shown on the short arm of chromosome 6, including the class I, II and III regions. A few key genes relevant for this thesis are shown, their location is according to National Center for Biotechnology Information's build 37 (hg19). For a complete MHC map see Horton *et al.* (61).

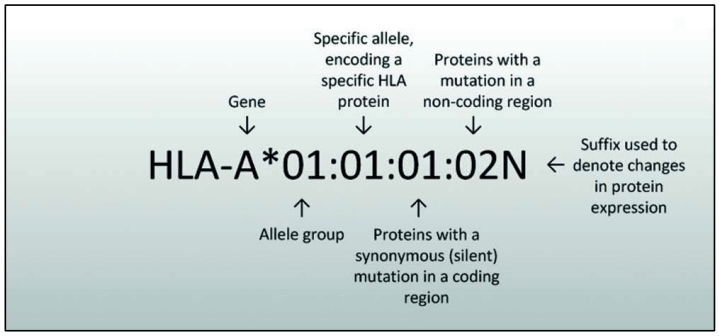

Figure 7: Nomenclature of an human leukocyte antigen allele.

The figure is modified from http://hla.alleles.org/nomenclature/naming.html

The function of HLA class I and class II molecules is generally considered separately, although these most likely share a common evolutionary history (62). Both sets of proteins are located on the cell surface where they present peptides to receptors on T cells. HLA class I molecules, e.g. HLA-A, HLA-B and HLA-C, are expressed on all nucleated cells and present intracellular peptides to CD8+ T cells, whereas HLA class II molecules, e.g. HLA-DR, HLA-DP and HLA-DQ, are expressed on antigen presenting cells (e.g. macrophages, dendritic cells, B cells) and present extracellular peptides to CD4+ T cells. HLA class I molecules are also

shown to act as ligands for receptors mainly found on natural killer (NK) cells, so-called killer immunoglobulin-like receptors (KIRs) (63), as well as receptors on cells of monocyte lineages; leukocyte immunoglobulin-like receptors (64). The function of HLA class I and II molecules is illustrated in Figure 8.

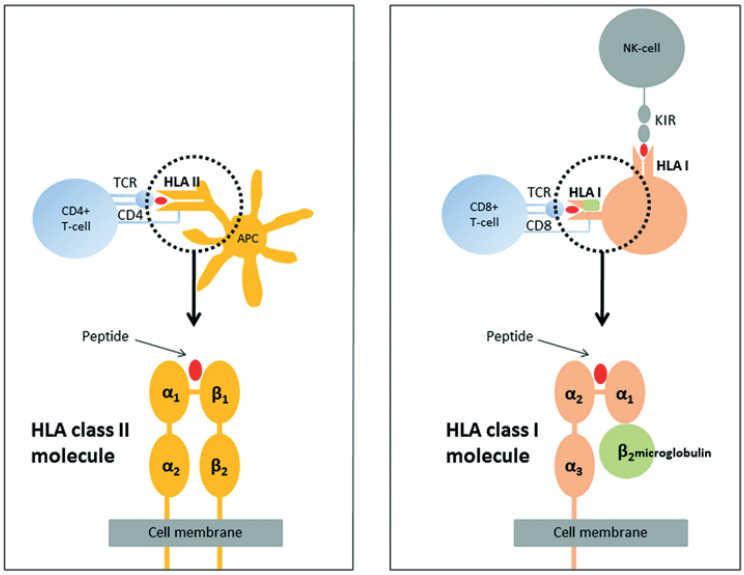

Figure 8: Function of classical human leukocyte antigen class I and II molecules.
Human leukocyte antigen (HLA) class II molecules (left figure box) are transmembrane glycoproteins, consisting of α and β polypeptide chains. The α-chain is encoded by classical HLA class II genes with an "A" in its name, e.g. *HLA-DRA, HLA-DQA1, HLA-DPA1*. Similarly, the β-chain is encoded by classical HLA class II genes with a "B" in its name, e.g. *HLA-DRB, HLA-DQB1, HLA-DPB1*. HLA class II molecules are found on antigen presenting cells (APC), and present peptides from outside the cell to CD4+ T cells (T helper cells). HLA class I molecules (right figure box) are also transmembrane glycoproteins, however the α-chain, encoded by classical HLA class I genes (i.e. *HLA-A, B, C*), is linked to a smaller polypeptide, called β_2-microglobulin, which is encoded on a different chromosome (chromosome 15). HLA class I molecules are found on all nucleated cells and present peptides from inside the cell to CD8+ T cells (cytotoxic T cells). As illustrated in the figure, HLA class I molecules can also act as ligands for receptors mainly found on natural killer (NK) cells, so-called killer immunoglobulin-like receptors (KIRs). TCR: T cell receptor.

The MHC is the most intensively studied genetic region in the human genome mainly due to the classical HLA genes, reflecting their importance and relevance in organ transplantation as well as immune mediated and infectious diseases. Although genetic associations within the MHC have been reported for numerous of immune related conditions, the complexity of this region represents a tremendous challenge for scientists working within the field of MHC

genetics. There are mainly four characteristics of this genetic region that pose problems for MHC geneticists. The first is the complex and variable LD pattern of this region. Although the generally high degree of LD allows for efficient and accurate genotype imputation in this region (as done in paper I), it also poses a great challenge. Due to the high degree of LD, highly conserved haplotypes exist, which can, at the most, span the majority of the MHC, stretching all the way from class I to class II. This means that if association with an extended haplotype encompassing hundreds of genes is shown for a certain condition, it is difficult, or even impossible, to pinpoint which part (e.g. gene) of the haplotype that represent the actual causal locus/loci. The high gene density, clustering of genes with related immunological function and the high variability of several of the genes in the MHC represent the second, third and fourth challenge. The latter is represented by the enormous polymorphism of several of the classical HLA genes and is mainly attributed to the peptide-binding groove of the HLA molecule, reflecting their ability to bind and present various peptides. However variation in non-coding regulatory regions affecting gene expression and alternative splicing of proteins, as well as interaction (i.e. epistasis) between HLA molecules and related molecules, are likely to add another layer to the polymorphism observed for the classical HLA molecules.

1.3.4 HLA molecules and solid organ transplantation

The first human leukocyte antigen,"MAC" (later named HLA-A2), was described in 1958 by Jean Dausset who discovered that patients who received blood transfusions formed antibodies against antigens on transfused leukocytes (65). Skin and kidney graft experiments in the 1960s demonstrated that human leukocyte antigens are strong histocompatibility antigens (reviewed in (66)), and attempts to transplant organs or tissue between humans had been unsuccessful prior to these discoveries.

The HLA antigens are, together with the ABO blood group system, the most important transplantation antigens. If these antigens differ in recipient and donor, the immune system of the recipient will recognize antigens on the transplanted graft as foreign, induce an immune response, and subsequent dysfunction or failure of the transplanted graft, i.e. graft rejection, either acute or chronic, may occur. Although post transplantation survival has greatly improved due to development of immunosuppressive agents to counteract acute and chronic rejection of the transplanted graft, rejection still represents a major reason for graft failure. Donor-recipient HLA matching is done prior to transplantation of most organs, e.g. kidney and heart, as HLA compatibility between donor and recipient has proven to reduce the incidence of graft rejection and to increase graft survival (reviewed in (67, 68)). In contrast,

the benefit of HLA matching remains controversial in liver transplantation (reviewed in (69)), and is per date not normally done. The reported incidence of acute rejection in liver transplanted patients varies (ranges from 6-80 %) (70, 71), however acute rejection is observed more often in patients with PSC (estimated to occur in ~50%) than other conditions listed for liver transplantation (reviewed in (25)). This has lead to the speculation as to whether there is a common link between PSC and acute rejection. Since PSC is strongly associated with genetic variants within the HLA complex, the HLA genotype of the recipient can be hypothesized to influence the risk of acute rejection of the transplanted graft. In addition, as HLA class I molecules can interact with KIRs on NK cells, donor HLA and recipient KIR genotype may affect the activation status of NK cells, and hence the alloreactivity mediated by NK cells in acute rejection. In fact, correlation between donor HLA-C KIR-ligand and graft survival has been reported in liver transplanted patients (72), yet another report failed to replicate this (73). Due to the conflicting results mentioned above, the correlation between HLA and KIR genotypes and risk of acute rejection were assessed in 143 donor-recipient pairs listed for liver transplantation in paper III, with particular emphasis on PSC patients.

1.3.5 The MHC and disease associations

The MHC is associated with more diseases (mainly autoimmune, immune mediated and infectious) than any other region of the genome (74). For many of these diseases, the statistical significance of the MHC association signal is often superior, dwarfing other genetic association signals in the genome in comparison. Many MHC associations were initially found by typing (first by serological typing, later by sequenced-based typing) of classical HLA class I and II genes. More recently, cost- and time effective indirect typing (i.e. imputation), of classical HLA alleles (and their amino acid variation), based on high density SNP genotypes, has been possible (75, 76). This has facilitated analyses of classical HLA alleles alongside MHC SNP genotypes in large cohorts, making it possible to fine map disease associations in the MHC.

Table 2: Association with classical human leukocyte antigen (HLA) class I and II loci and alleles in various immune mediated diseases.

Disease	*HLA-A*	*HLA-C*	*HLA-B*	*HLA-DRB1*	*HLA-DQA1*	*HLA-DQB1*	*HLA-DPA1*	*HLA-DPB1*
PSC (1, 2, 77-87)	A*01	C*07	B*07 B*08	DRB1*03:01 DRB1*04 DRB1*07:01 DRB1*11 DRB1*13:01 DRB1*15:01	DQA1*01:02 DQA1*01:03 DQA1*02:01 DQA1*03 DQA1*05:01	DQB1*02:01 DQB1*03:01 DQB1*03:02 DQB1*03:03 DQB1*06:02 DQB1*06:03		
type 1 diabetes (88, 89)	A*01 A*11 A*24		B*13 B*18 B*39 B*50	DRB1*03:01 DRB1*04	DQA1*03:01 DQA1*05:01	DQB1*02:01 DQB1*03:02		rs6457721
coeliac disease (90)					DQA1*02:01 DQA1*03 DQA1*05:01	DQB1*02:01 DQB1*02:02 DQB1*03:02		
primary biliary cirrhosis (91, 92)			B15 B41 B55 B58	DRB1*02 DRB1*08 DRB1*11 DRB1*13 DRB1*14	DQA1*04:01	DQB1*04		DPB1*03:01
ulcerative colitis (93)	A*02:01 A*29:02	C*03:04 C*07:02 C*12:02 C*16:02	B*07:02 B*18:01 B*44:03 B*52:01	DRB1*01:03 DRB1*03:01 DRB1*04 DRB1*07:01 DRB1*09:01 DRB1*11:01 DRB1*13:01 DRB1*13:02 DRB1*15:01 DRB1*15:02	DQA1*01:01 DQA1*02:01 DQA1*03:01	DQB1*02:01 DQB1*02:02 DQB1*03:02 DQB1*03:03 DQB1*03:03 DQB1*03:01 DQB1*05:01 DQB1*06:01 DQB1*06:02 DQB1*06:03 DQB1*06:04		DPB1*03:01 DPB1:04:01 DPB1*06:01 DPB1*11:01
Crohn's disease (93)	A*03:01	C*04:01 C*06:02 C*07:01 C*08:02 C*12:02 C*14:02	B*08:01 B*14:02 B*35:02 B*35:03 B*52:01 B*57:01	DRB1*01:01 DRB1*01:03 DRB1*03:01 DRB1*07:01 DRB1*08:01 DRB1*13:02 DRB1*16:01	DQA1*04:01	DQB1*02:01 DQB1*04:02 DQB1*05:02 DQB1*06:09	DPA1*01:03	
rheumatoid arthritis (94-96)			Amino acid position 9	DRB1*01 DRB1*03:01 DRB1*04 DRB1*07:01 DRB1*10 Amino acid position 11, 71 and 74				Amino acid position 9
multiple sclerosis (97)	A*02:01			DRB1*03:01 DRB1*13:03 DRB1*15:01		DQB1*02:01		
ankylosing spondylitis (98)			B14:03 B27 B39 B60				DPA1*01:03 DPA1*02:01	DPB1*11:01 DPB1*13:01
myastenia gravis (99, 100)	A*01	C*07:01	B*07 B*08	DRB1*03:01 DRB1*07:01 DRB1*13:01 DRB1*15:01 DRB1*16		DQB1:02:01		
narcolepsy (101, 102)				DRB1*03:01 DRB1*15:01 DRB1*13:01	DQA1*01:02	DQB1*02:01 DQB1*06:02 DQB1*06:03		
psoriatic arthritis (103)		C*06:02	B13 B27 B38/39 B57	DRB1*04				

With a few exceptions, various diseases associate most strongly with alleles of classical HLA class I and II genes, as opposed to non-HLA genes within the MHC. The number and collection of associated HLA alleles, as well as their genetic localization (i.e. class I or II), vary between diseases, and are often a mix of both protective (i.e. negatively associated) and risk (i.e. positively associated) alleles. However, as Table 2 demonstrates, similarities also exist as several conditions are associated with identical HLA alleles, e.g. DRB1*15:01, indicating MHC pleiotropy (i.e. genetic overlap), between various diseases. Only for a few conditions has it been possible to narrow down the MHC association to a primary causative loci/allele, as observed for the *HLA-DQ* locus in coeliac disease (90), and the occurrence of the HLA-DQA1*01:02-DQB1*06:02 haplotype and the HLA-B*27 allele in almost all narcolepsy patients (90 %) (101) and ankylosing spondylitis patients (80-95 %) (98, 104), respectively. Most MHC associated conditions exert a more complex picture due to association with several haplotypes, many of these of considerable length due to extensive LD, as for instance association with the ancestral haplotype 8.1 (AH8.1; A*01-B*08-C*07-DRB1*03:01-DQA1*05:01-DQB1*02:01) observed in PSC, type 1 diabetes, Crohn's disease and myasthenia gravis. Another possible contribution to the complexity is the heterogeneous nature of many MHC associated diseases. Recent years, clinically distinct subgroups of several MHC associated diseases have shown to have different MHC associations, as observed in rheumatoid arthritis with and without autoantibodies (105). And so, it still remains a challenge for many MHC associated diseases (and their potential subgroups) to interpret the genetic complexity and firmly establish the underlying MHC architecture. Once this has been done, the search for the functional relevance of the various associations can be targeted.

1.3.6 MHC and PSC

When I started my PhD, many classical HLA class I and II genes and alleles had been found to be associated with either risk or protection of PSC. Genotyping and comparison of classical HLA alleles in several candidate driven case-control studies of variable sizes, mostly of European origin, had been done. In addition, the superior role of the MHC in PSC genetic architecture had been firmly established in two PSC GWAS (40, 41).

The first HLA alleles found to be associated with PSC was the HLA class I allele, HLA-B*08 and the HLA class II allele, HLA-DRB1*03:01, both reported in the early 1980s (1, 2). Although located far apart, these alleles are found on the same extended haplotype stretching from class I to class II, namely the AH8.1 (Figure 9). Association with alleles on AH8.1 is the

most prominent MHC association in PSC and has been replicated several times (1, 2, 77-79, 81-87). However, teasing apart the causative variant(s) on this haplotype has proven difficult due to its extensive LD, and other alleles of the AH8.1, but B*08 and DRB1*03:01, have been reported. This includes HLA-C*07, DRB3*01:01, DQA1*05:01, DQB1*02:01, in addition to alleles of non-classical HLA genes, i.e. the major histocompatibility complex class I chain-related A (MICA) *008/5.1 allele (87, 106, 107) and the TNFα promoter -308 A allele (108, 109). Interestingly, fairly recent studies have indicated a potential primary effect of the class I region of the AH8.1 haplotype in PSC. Association with B*08 and not DRB1*03:01 was found in an African American PSC population (with lesser degree of LD) (110). In addition, the top MHC association signal resided in the vicinity of *HLA-B* in both PSC GWAS (40, 41).

HLA class II genes have been the focus of several PSC studies. In addition to the class II alleles on AH8.1, risk associations have been reported with haplotypes carrying DRB1*13:01 and DRB1*15:01 (77, 78, 80, 84-86). DRB1*15:01 is located on yet another extended haplotype; i.e. the ancestral haplotype 7.1 (AH7.1), including, among other alleles, HLA-C*07 and HLA-B*07 (Figure 9). The extension of the DRB1*13:01 haplotype is less pronounced and largely confined to the class II region. Protective associations of HLA class II haplotypes carrying HLA-DRB1*04, HLA-DRB1*07:01 and HLA-DRB1*11 have been reported (78, 80, 85, 86).

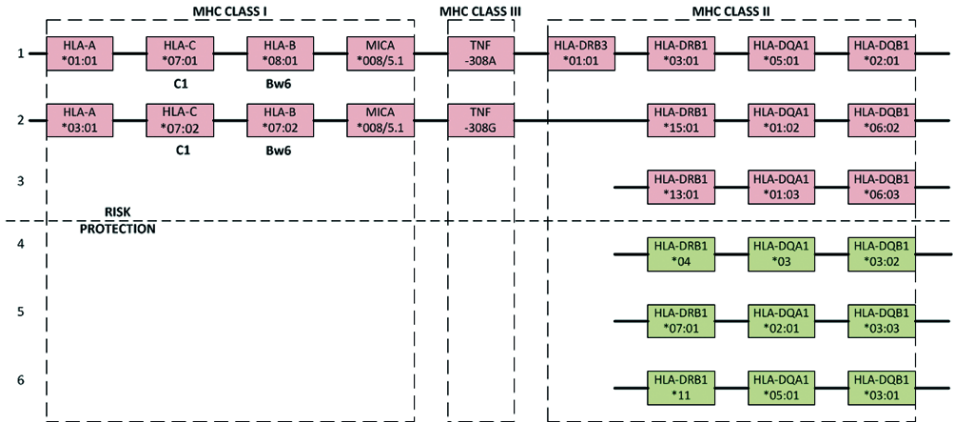

Figure 9: Reported major histocompatibility complex haplotype associations in primary sclerosing cholangitis.

Six haplotypes shown to be associated with primary sclerosing cholangitis (PSC) prior to presented work in this thesis. Haplotype 1, 2 and 3 are associated with risk of PSC, while haplotype 4, 5 and 6 are associated with protection. Haplotype 1 and 2 are known as ancestral haplotype 8.1 (AH.81) and ancestral haplotype 7.1 (AH7.1), respectively. C1 and Bw6 below *HLA-C* and *HLA-B* at haplotype 1 and 2 refer to classification of the *HLA-C* and *B* alleles according to their killer immunoglobulin-like receptor (KIR) ligand group. HLA: human leukocyte antigen. MHC: major histocompatibility complex. MICA: MHC class I chain-related A. TNF: tumor necrosis factor. The figure is modified from the doctoral thesis of Johannes E.R. Hov (111).

A potential functional relevance of the HLA class I association in addition to their peptide presenting ability to T cells, is the interaction between HLA class I molecules (predominantly HLA-B and HLA-C) and receptors on NK cells; KIRs. HLA class I molecules act as inhibitory ligands on KIRs and hence affect the activation status of NK cells, however the inhibitory potential of different HLA-KIR interactions varies. *HLA-B* and *HLA-C* alleles are grouped into Bw4/Bw6 and C1/C2 KIR ligand groups, respectively. Ligand groups with stronger inhibitory potential, i.e. Bw4 and C2, were found to protect against PSC in a Scandinavian population (106). Replication of the protective role of C2, but not Bw4, was found in Italian PSC cohort (80).

Little has been done in order to translate the genetic associations with the MHC into functional relevance for PSC, except one study. Given the significance of *HLA-DRB1* associations in PSC, amino acid variation of the HLA-DRβ molecule was investigated in a Scandinavian cohort (112). This resulted in two disease associated residues (i.e. residue 37

and 86), one of which (i.e. residue 37) influence the electrostatic properties of pocket P9 in the peptide binding groove of the HLA-DR molecule. Both the PSC associated *HLA-DRB1* risk alleles HLA-DRB1*03:01 and HLA-DRB1*13:01 encode Asparagine at residue 37 which induces a positive charge in pocket P9. This finding support the assumption that disease associated HLA molecules are able to present a different range of peptides than non-associated molecules and thereby contributing to a possible disease causing immune response.

To sum up, at the beginning of my PhD, the MHC association in PSC was established as the major genetic contribution to disease risk. However the complexity of the association was evident and many classical HLA alleles on several MHC haplotypes of variable extension had been reported. Further refinement of the MHC architecture in PSC was addressed in paper I and II.

2 AIMS OF THE PROJECT

The aim of studying the genetic predisposition to any condition is to unravel the molecular mechanisms causing disease. The MHC harbors the major genetic risk(s) for PSC, and thus represents the biggest and most important clue to PSC disease pathogenesis.

The main aim of this thesis was to continue previous work and to further characterize the genetic association within the MHC in PSC at chromosome 6p21 by:

- Fine-mapping the MHC association in PSC and search for potential primary effects on associated haplotypes (paper I).
- Investigate whether phenotypical diversity between small duct PSC and classical PSC can reflect differences in HLA risk (paper II).

Furthermore, PSC patients seem to be at greater risk of early rejection of the transplanted liver graft after liver transplantation, which is the only curative treatment of PSC as of date. Given the strong MHC association with PSC, another aim of this thesis was to:

- Investigate whether genetic variants of classical HLA class I and II molecules and KIRs (receptors for HLA class I), affect risk of rejection after liver transplantation (paper III).

3 MAIN RESULTS

3.1 Refinement of the MHC Risk Map in a Scandinavian Primary Sclerosing Cholangitis Population (paper I)

The aim of this study was to map the main determinants that constitute the MHC association signal in PSC. Three hundred and sixty five PSC patients and 368 healthy controls of Scandinavian ancestry were included. Various genotype data were incorporated in the statistical analyses, including sequencing-based classical HLA typing, SNP genotypes and previously reported determinants of the PSC associated MHC risk; i.e. amino acid variation of HLA-DRβ, the *MICA* transmembrane microsatellite polymorphism and *HLA-B* and *HLA-C* ligand properties for killer immunoglobulin-like receptors. A breakdown of the association signal was done by two statistical approaches, including (i) multivariate logistic regression modelling and (ii) univariate regression analysis followed by three subsequent conditional strategies. In summary, nine PSC associated MHC haplotypes of variable extension were identified by combining these two statistical approaches. Eight of these were tagged by classical class I or II HLA alleles, predominantly at *HLA-B* and *HLA-DRB1*. A novel independent association was detected near *NOTCH4* in the class III region, tagged by rs116212904 (OR = 2.3, 95 % CI:1.8-3.0, $P = 1.35 \times 10^{-11}$).

3.2 Small duct primary sclerosing cholangitis without inflammatory bowel disease is genetically different from large duct disease (paper II)

The main objective of this study was to characterize potential genetic similarities and/or differences between two phenotypically distinct subgroups of PSC, i.e. small duct PSC (~10 % of the overall PSC population) and large duct PSC (~90 % of the overall PSC population). This was done by investigating genetic associations at four key HLA loci (*HLA-A, HLA-B, HLA-C, HLA-DRB1*) within the MHC, which represent the strongest genetic risk factors in large duct PSC. Eighty seven small duct PSC patients, 485 large duct PSC patients and 1117 healthy controls from Scandinavia, UK and US were included. Only two out of the eight associated large duct PSC HLA alleles in this study were found to be significantly associated with small duct PSC; HLA-DRB1*13:01 (OR = 2.0, 95 % CI: 1.2-3.4, $P = 0.01$) and HLA-B*08 (OR = 1.6, 95 % CI: 1.1-2.4, $P = 0.02$). Further stratification of small duct PSC patients based on co-existing IBD diagnosis showed that small duct patients with co-existing IBD greatly resembled large duct disease in its HLA association. In total five (HLA-B*08, DRB1*13:01, C*07, DRB1*03 and DRB1*04) out of eight HLA alleles that associated with

large duct PSC were also associated with small duct PSC with IBD. In contrast, small duct PSC without co-existing IBD was only associated with DRB1*13:01. The genetic similarity between small duct PSC with co-existing IBD and large duct PSC, suggests that the former could represent early stage or a milder variant of large duct disease. Small duct PSC without IBD however, could represent cholangiopathies of other etiologies, due to the difference in HLA predisposition detected in this study.

3.3 HLA variants related to primary sclerosing cholangitis influence rejection after liver transplantation (paper III)

In this study influence of HLA and KIR genotypes on risk of acute rejection after liver transplantation was investigated. Hundred and forty three donor-recipient pairs listed for liver transplantation at Oslo University Hospital, Norway, between 1996 and 2008 were included. 45.5 % of the liver recipients had underlying immune-mediated disease (i.e. PSC (30.8 %), primary biliary cirrhosis (PBC) (9.8 %) and autoimmune hepatitis (AIH) (4.9 %)). Viral hepatitis and other liver conditions constituted 13.9 % and 43.9 % of the recipient population, respectively. Acute rejection was confirmed in 29 % of the overall recipient population, with a significantly higher frequency in patients with PSC compared with non-PSC patients (41 % *vs.* 23 %, OR = 2.3, 95 % CI: 1.1-4.9, $P = 0.03$). Donor-recipient matching for HLA class I and II (*HLA-A, B* and *DRB1*) had no significant impact on risk of acute rejection. Copy number variation (CNV) and presence of the various KIR genes did not affect risk of acute rejection, nor did KIR ligand (i.e. HLA-C and Bw4) disparity between donor and recipient or the donor KIR ligand genotype. An increased risk of acute rejection was found for three PSC associated HLA risk alleles in the overall recipient population, i.e. HLA-B*08 (OR = 2.5, 95 % CI: 1.4-4.6, $P = 0.002$), HLA-C*07 (OR = 2.5, 95 % CI: 1.4-4.1, $P = 0.001$) and DRB1*03 (OR = 1.9, 95 % CI: 1.0-3.3, $P = 0.04$). A protective effect on acute rejection was observed for the DRB1*04 allele (OR = 0.2, 95 % CI: 0.1-0.5, $P = 0.0005$), also known to be protective of PSC. These associations remained evident (except for DRB1*03) in a subgroup analysis of non-PSC recipients, indicating that HLA variants of recipients influence risk of acute rejection after liver transplantation, irrespective of underlying liver disease.

4 METHODOLOGICAL CONSIDERATIONS

4.1 Study design and study population

All studies included in this thesis have a population based case-control study design. This is the method of choice when the objective is to study associations between genotype and complex traits (113).

Several properties of the study population, represented by cases and controls, affect the chance of detecting true genetic associations. First, it is important to define the phenotypic trait under study precisely, as a homogeneous disease population increases the power to detect true genetic associations and is less exposed to spurious genetic findings. To accommodate this, standard diagnostic criteria for the diagnosis of PSC (19), small duct PSC (114), acute rejection (115) and IBD (116) were employed in the respective studies in this thesis. Nevertheless, the phenotypic diverse nature of PSC is present in our study populations, represented by for instance variations in progression of disease and presence of co-morbidities, and might reflect genetic variations. An additional challenge in paper II is the likely "contamination" of true large duct PSC patients in our small duct PSC population, as ~25 % of small duct PSC patients are known to progress to large duct PSC with time. To estimate a percentage of such possible "contaminating" large duct PSC patients, Monte Carlo simulations, calculated in MATLAB, were performed based on observed frequencies of the large duct PSC associated risk allele, HLA-B*08, in our small duct population compared with frequencies observed in large duct PSC and healthy controls. This resulted in an estimated 32 % (95 % CI: 4-65 %) large duct PSC individuals in our small duct PSC study population. Stratification of the disease population based on phenotypic differences (e.g. IBD *vs.* non-IBD) will increase its homogeneity and may lead to genetic enrichment (as done in paper II and III). However, this causes loss of power and increased risk of type II errors due to smaller study populations. The majority of patients were recruited at highly specialized referral centers, indicating that these patients may be more severely affected and perhaps not representative for the disease group in general. Yet, this may also be an advantage as this group is more homogeneous and thus empowered to detect genetic associations (117).

Equally important to define the phenotypic trait under study precisely, is to assure a representative control population, i.e. they should represent the population at risk of disease (118). In an ideal genetic association study cases and controls should be comparable in age, ethnicity, gender, occupation and habits, and preferably screened for diseases that may

influence the results. This is however challenging, both cost- and time-wise, and practical adjustments must be done in order to establish the best possible control population. Various sources of control populations have been used in the present studies; healthy bone-marrow donors (paper I and II), healthy volunteers (paper II), generated controls (based on non-transmitted alleles from parents of type 1 diabetic sib-pairs (119), paper II) and deceased liver donors (paper III). As PSC has a male predominance, efforts have been made to match for gender in our control population. Age matching and screening for PSC in controls have not been done in the present or previously published genetic studies of PSC. Due to the low prevalence of PSC (~1/10.000), presence of undiagnosed cases among our controls is negligible and will not affect the power to detect genetic associations (120). Cases and controls must also be matched according to their ethnic origin to avoid systematic differences in allele frequencies observed between populations of different ethnicity, referred to as population stratification (121). In paper I and III only individuals of Scandinavian ancestry were included, and there was no need to adjust for population structure. This is supported by previous studies in which no significant population stratification was found for this population (41, 45). In paper II cases of additional geographical origin were included (i.e. UK and US) and geographically matched controls were included in order to minimize the possible confounding effect of population stratification. However, although only individuals of self-reported Caucasian decent were included from these countries, a potential minor effect of population stratification in paper II is possible as especially the Caucasian US population reflect a more heterogeneous genetic background than Caucasians of Scandinavian ancestry.

Table 3: Overview of cases and controls included in the various studies.

Panel	**Paper I**		**Paper II**			**Paper III**		
	PSC	Controls	Small duct PSC	PSC	Controls	Liver Recipients		Liver donors
						PSC	Non-PSC	
Scandinavia	365	368	43	357	368	44	99	143
United Kingdom	-	-	28	77	600	-	-	-
United States	-	-	16	51	149	-	-	-
Total	**365**	**368**	**87**	**485**	**1117**	**143**		**143**

4.2 Sample preparation

DNA extracted from peripheral blood was used in this thesis. DNA was extracted over a long time period and in multiple laboratories involving various DNA extraction protocols. Diversity of DNA extraction procedures does not seem to influence the yield and purity of DNA produced at a great extent (122). However, storage and processing of blood and DNA, such as repeated freeze-thaw cycles, may influence the quality of DNA (122). For practical purposes, low quality DNA will fail to generate genotypes successfully, and hence the genotyping success rate is of importance. This and other post-genotyping quality control (QC) steps are discussed in section 4.4. To minimize the requirement of genomic DNA, whole-genome amplified (WGA) DNA was used to genotype SNPs and HLA alleles. Although WGA DNA is suitable for SNP genotyping arrays and HLA sequencing, it may lead to imbalanced amplification of the genome and thus a potential source of bias when determining CNVs in the genome (123). Genotyping of KIR genes, and their CNVs, as performed in paper III, was done with genomic DNA as input.

4.3 Genotyping of HLA alleles, SNPs, KIR genes and the MICA 5.1 polymorphism

4.3.1 HLA genotyping

HLA typing in this project has exclusively (paper I and III) or mostly (paper II) been performed at Institute of Immunology, Oslo University Hospital, Rikshospitalet. Primarily, the method of sequence-based typing was performed using "in-house" protocols, involving three essential steps; i) PCR amplification of specific coding regions of HLA genes, i.e. exon 2 and 3 for HLA class 1 loci, and exon 2 for HLA class 2 loci, ii) sequencing of the amplicons and iii) assignment of alleles. Assign Sequence Based Typing Software (Conexio Genomics) was used to assign HLA alleles correctly (step 3). It compares sequence results with the IMGT/HLA database (http://www.ebi.ac.uk/ipd/imgt/hla/) to generate possible allelic combinations. However, sometimes so-called ambiguous allele combinations arise, i.e. two (or more) different allele combinations yield identical nucleotide sequence (124). Multiple actions can be taken to address the issue of ambiguous combinations, of which several were employed in this thesis; (a) down-scaling results to a 2-digit resolution, (b) retyping samples with sequence-specific oligonucleotides or sequence-specific primers, and (c) selecting the most probable allele combination based on allele frequencies reported in the ethnic group matching your study population (a strategy that may lead to underestimation of rare alleles).

Some HLA class I alleles have identical sequences across exon 2 and 3, for these it is necessary to sequence additional exons to obtain higher than 2-digit resolution. In all the papers included in this thesis, HLA class I alleles (*HLA-A, B* and *C*) were resolved to a 2-digit resolution and HLA class II alleles (*HLA-DRB3, DRB1* and *DQB1*) to a 4-digit resolution. For *HLA-A* and *HLA-B* alleles it was necessary to recode the alleles to serotype nomenclature in patients (paper I and II), to ensure comparability with the Scandinavian control population, for which serotyping based methods were applied at these two loci. To reduce the number of comparisons in statistical analysis *HLA-A* and *HLA-B* alleles were also reported at serotype level in paper III. For the same reason, *HLA-DRB1* alleles were reported at a two digit resolution both in paper II and III. HLA-C and HLA-Bw KIR ligand groups were determined in paper I. A dimorphism of the amino acid sequence of HLA-C at position 80 defines the two distinct HLA-C KIR ligand groups; C1 (Asn80) and C2 (Lys80) (125). *HLA-C* alleles were allocated to C1 or C2 based on their nucleotide sequence at exon 2. For HLA-B, amino acids at position 77-83 subgroup *HLA-B* alleles into Bw4 and Bw6 (126), where only Bw4 act as a KIR ligand. Bw4 and Bw6 specificities were deduced from the *HLA-B* alleles based on known sequence information (127).

4.3.2 SNP genotyping

Two different types of SNP genotyping technologies were applied in paper I.

The SNPlex Genotyping system (Applied Biosystems) was used to genotype SNPs spanning the MHC in paper I. This multiplex genotyping system had the ability to genotype 48 different SNPs simultaneously and was an efficient way of genotyping multiple SNPs when the genotyping was initiated in 2005. Although this technology today has been replaced by more efficient SNP arrays including hundreds of thousands of SNPs, it was well validated and showed high concordance with other genotyping technologies, such as TaqMan® genotyping (128). In total 10 SNPlex assays (i.e. 480 SNPs) were genotyped. In order to assure coverage of SNPs throughout the MHC, 432 tag SNPs from the Celera database (https://www.celera.com/) were selected. Additional 48 SNPs were selected based on their reported association with other immune mediated diseases.

Replication genotyping of rs116212904 in paper I was done by TaqMan® technology (Applied Biosystems). This technology is well validated and suitable for genotyping a few number of SNPs (129). Replication genotyping should preferably be done by a different technology in a different study population (130). Therefore, in an attempt to strengthen the

validity of our SNP association, an alternative genotyping technology, i.e. TaqMan®, was applied.

4.3.3 KIR genotyping

The KIR genes are encoded on chromosome 19q13.4. These encode activating and inhibiting receptors found on NK cells and some T cells (125). This genetic region shows, much like the MHC, great variability between individuals, both in genetic content (i.e. number of KIR genes), allelic diversity and CNV of the various genes. In order to explore diversity beyond genetic content (i.e. presence/absence), 14 KIR genes and 2 pseudogenes were typed for CNV using a quantitative PCR method (131). The protocol was developed at Cambridge Institute for Medical Research, University of Cambridge, where also the typing was performed (132). Each assay allowed for detection of 2 KIR genes and 1 reference gene (of known copy number (=2)). Two individuals with known copy number for each KIR gene, and a non-template control, were included in each run. The primers and probes were designed specifically in order to avoid known allelic diversity within each gene, so that annealing to, and amplification of all possible alleles, were done. However, it remains a possibility that unknown alleles with rare polymorphisms at the annealing site for primer or probe exist in our data set, which may disrupt amplification and consequently cause incorrect assignment of copy number(s) for the KIR gene in question. Each sample was typed in quadruplicate in order to assure accurate copy number scoring.

4.3.4 Genotyping the MICA 5.1 polymorphism

Presence of the previously reported PSC associated MICA 5.1 polymorphism was included in paper I. Genotyping results were obtained from a previously published study (106), in which fragment length determination was used to separate the various *MICA* microsatellite polymorphisms, as determined by number of trinucleotide (GCT) repeats in the transmembrane region. Amplified products were separated on a gel, and identified according to size on a DNA sequencer, as described in (87).

4.4 Post genotyping quality control

Following efforts to recruit an unbiased study population, as described in section 4.1, and subsequent genotyping, it is important perform QC of genotype data to ensure high quality data sets prior to statistical association testing.

4.4.1 HLA data

Several factors may influence the genotyping success rate, including DNA quality and assay specific factors, e.g. appropriate primer and probe design. HLA typing success rate, excluding unsuccessful genotyping results and unresolved ambiguities, in the various papers is presented in Table 4.

Table 4: Human leukocyte antigen typing success rate in the various studies.

	Paper I Success rate	**Paper II** Success rate	**Paper III** Success rate
HLA-A	0.99	0.99	0.993
HLA-B	0.98	0.98	0.993
HLA-C	0.97	0.98	0.993
HLA-DRB3	0.995	-	-
HLA-DRB1	0.997	0.99	0.993
HLA-DQB1	0.98	-	-

4.4.2 SNP data

QC of SNP genotype data may involve several steps to filter out low-quality markers and origins from protocols established for genome-wide association data (133). The first step is to call genotypes, this is typically done by calling algorithms. For studies genotyping a smaller set of SNPs, additional assessment of correct genotype assignment can be done by visual inspection of so-called cluster plots. Failure to assign genotypes properly is indicated by poorly defined clusters. Figure 10 demonstrates a cluster plot from the TaqMan® analysis in paper I.

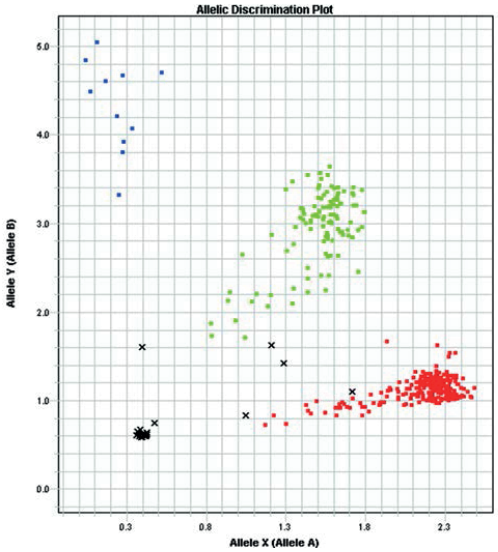

Figure 10: Cluster plot

Blue squares indicate individuals with two copies of allele B, red squares indicate individuals with two copies of allele A and green squares indicate individuals with one copy of each. Undetermined samples are indicated with a black cross (many of which represented empty wells).

Visual inspection of cluster plots was done for all SNPs genotyped in paper I followed by several QC steps;

1) Identification of SNPs with excessive missing genotypes across samples tested. This is important, as it is indicative of low quality genotyping for the SNP in question. Classically, SNPs with a genotyping call rate of < 95 % is excluded for further analysis (133). Accordingly, a threshold of 95 % was set in our study (paper I).
2) SNPs deviating from Hardy-Weinberg Equilibrium (HWE) (*P*-value<1 x 10^{-4}) in controls were excluded. The Hardy-Weinberg law assumes that in a large population with random mating and no selection, migration or mutation, the genotype frequencies will be in equilibrium from generation to generation. Deviation from HWE can indicate population stratification, inbreeding, as well as problems with genotyping (e.g. contamination of samples) or genotype-calling (133, 134). A SNP which is strongly associated with disease may deviate from HWE and it is therefore recommended that testing for deviation from HWE should only be done in controls (133).

3) Finally, SNPs with a minor allele frequency (MAF) of <1 % were excluded. The small sizes of the heterozygote and rare homozygote genotypes, make these variants difficult to call by calling algorithms, and they frequently present as false positives in case-control association tests.

In total 405 SNPs passed QC, with a final genotyping rate of 0.99 in our study population. Replication genotyping of rs116212904 by TaqMan® achieved a genotyping success rate of 0.98.

4.4.3 KIR data

Post genotyping QC of KIR typing data included accurate calling of copy numbers. Calling was done by CopyCaller v.1.0 (Applied Biosystems). Several actions were taken to ensure accurate scoring of copy numbers; a) Visual inspection in CopyCaller was done for each KIR gene in all individuals. If intermediate copy numbers were present, results were discussed based on known KIR haplotype architecture, and if inconclusive, excluded. b) Samples with ≤ 1 successful replicate, or with > 4 copy numbers, were excluded. c) All 0 copies were verified by amplification of the reference gene, in order to exclude unsuccessful genotyping. It is a possibility that these actions (a and b) may have caused incorrect exclusion of rare haplotypes of unusual gene content, not previously described. Genotyping success rate for the KIR genes and pseudogenes ranged from 0.99-1.00.

4.5 Imputation of SNP genotypes

Genotype imputation is a statistical method to infer genotypes not directly typed in your study population, based on haplotype patterns in a more densely genotyped reference panel. By increasing the number of SNPs available for association testing, a boost in power can be achieved by imputation algorithms. It can also facilitate fine mapping of an associated region, increasing the possibility of identifying causal variants. Several imputation methods are available, including MaCH, Beagle, and IMPUTE, these have shown to produce comparable results, all with error rates between 5-6 % (135). MaCH 1.0 (136) and minimac v. 2012.5.29 (137, 138) were used in paper I. Figure 11 shows how imputation works.

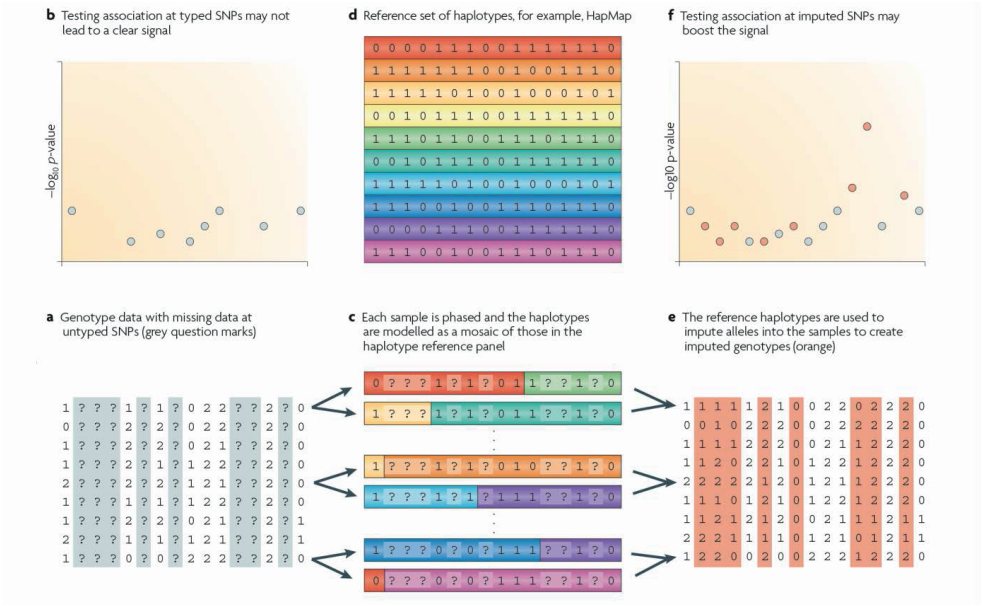

Figure 11: Illustration of the method of genotype imputation.

A reference set of haplotypes (part d), for instance from the 1000 Genomes project, is used to infer missing data (indicated by question marks in part a) and c) by mathematical algorithms. This indirect typing method is referred to as "imputation". Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Genetics (135), copyright (2010).

Imputation of SNP genotypes was done in paper I based on reference data from the 1000 Genomes Project (http://www.1000genomes.org/). Our input for imputation, 405 SNPs, was rather small compared to current standard and represent a limitation to our imputation performance. That being said, the extensive LD pattern within the MHC makes this region highly amenable to effective genotype imputation, even with a limited amount of SNPs as input. There are several factors that influence the quality of imputed genotypes. Firstly, genotyped SNPs that serve as input for imputation must be of good quality, i.e. they must pass post genotyping QC measures (QC measures applied in paper I are described in section 4.4.2). Secondly, the ethnicity of the reference panel should match the ethnicity of your study population. E.g. imputing SNP genotypes in our Scandinavian study population based on genotypes in an African reference panel would be wrong due to population stratification. However, a too restricted reference panel (e.g. only Scandinavians in our case) might cause lower imputation accuracy for low frequency alleles than if a more diverse reference panel is

included, due to unexpected allele sharing among populations (139). Therefore, all haplotypes of European ancestry (n = 758) at time of analysis, were used as reference panel in paper I. Thirdly, genotyping quality of SNPs in the reference panel is equally important as the quality of SNPs that serve as input and potential genotyping errors in the 1000 Genomes reference panel represent a source of bias to our imputation output. Lastly, quality of imputed SNPs must be assessed. We excluded all imputed SNPs with a MAF of < 1 %, because rare alleles are difficult to impute. In addition accuracy of imputed SNPs was assessed by the squared correlation (r^2) between predicted (i.e. imputed) and true genotypes. A MAF-specific r^2 threshold was employed to give an average r^2 of > 0.8. In the end, 18,644 SNPs were left for analyses, boosting the number of SNPs by 18,239.

4.6 Immunostaining for NK, B and T cells

Immunohistochemistry is the process by which antigens in cells are detected by their interaction with antibodies. This method can be used to visualize and quantify cell subsets in various tissues. Visual detection of the antigen-antibody interaction can be done by chemically linking the antibodies to an enzyme (e.g. peroxidase) which catalyzes a chemical reaction resulting in a colored product. Alternatively, the antibody can be detected using a fluorescent label (i.e. immunofluorescence). In paper III, immunostaining of formalin fixed liver allograft biopsies was done in order to detect the number and distribution of NK, B -and T cells in livers affected by acute rejection. Double staining was necessary to distinguish NK cells from natural killer T cells. Immunostaining was done by collaborators at Department of Pathology, Oslo University Hospital, Oslo, Norway.

4.7 Statistical methods

4.7.1 Genetic association testing

Univariate analyses

A genetic association case-control study compares the frequency of alleles or genotypes at genetic marker loci, e.g. SNPs or a gene, between individuals with (i.e. cases) and without (i.e. controls) a specific phenotype, e.g. a disease. Several genetic markers were tested for association with PSC (and acute rejection following liver transplantation) in this thesis.

Frequencies of HLA alleles and HLA genotypes, and presence of KIR genes, were tested for association by Chi-square test and Fisher's exact test (if necessary) in paper II and III. Association testing of HLA in paper II and III was mainly done by allele based tests. Due to

the large number of genotype categories for the multiallelic HLA loci, comparisons based on the presence or absence of an allele were originally established as the standard approach for HLA association studies (140). A general layout of a 2 x 2 table for a conventional HLA-disease association study is illustrated in Table 5. Odds ratios (ORs) and corresponding 95 % confidence interval (CI) were calculated by Woolf's formula with Haldane's correction (140).

Table 5: A 2 x 2 contingency table for association testing of an individual HLA allele

		Allele X	
		Present	Absent
Outcome	Patients	a	b
	Controls	c	d

Univariate logistic regression was used in paper I to test for association of various genetic markers within the MHC. This method is suitable when analyzing dosage data (as obtained from imputation algorithms), it also allows for inclusion of covariates (141), making it possible to conduct conditional strategies. Continuous variables in paper III were analyzed by student's t test and one way analysis of variance (ANOVA), after normal distribution of the data was confirmed. Meta-analyses across geographical panels were performed in paper II, Woolf's test was applied to test for heterogeneity of the individual group ORs, and the Mantel-Haenszel method was used to calculate common ORs.

Multivariate analyses

Multivariate logistic regression models were used in paper I to study the collective contribution of various HLA loci and their alleles to the MHC association in PSC, with disease state (i.e. disease *vs.* non-disease) as the dichotomous outcome variable.

To assess the contribution of the six HLA loci included in the study, a stepwise logistic regression approach was used (142). A baseline model containing a specific locus, e.g. *HLA-B*, was compared to models adding each of the other five loci, one at a time, using a likelihood ratio test. This was done for all the six loci included.

To evaluate which allele(s) across all six loci tested, that could best predict the outcome (i.e. disease status), an automatic stepwise variable selection method, the stepAIC function in R (http://cran.r-project.org/web/packages/MASS/), was applied. This method selects in each step the covariate that, by exclusion or inclusion to the regression model, decreases the

Akaike Information Criterion (AIC) the most (the smaller the value of AIC, the better the model). The stepAIC was run on all HLA alleles present in our study population simultaneously, significant covariates were kept in the final model.

4.7.2 Statistical significance, multiple testing and power

In genetic association studies, multiple markers are typically tested for association. Due to the large number of statistical tests performed, chances are high that statistical significance is observed only due to chance. Given a statistical significance level of 0.05, as many as one in 20 tests will by chance be significant, i.e. false positive (type I error). Controlling for multiple testing is an important aspect of studies involving many genetic markers (141). An accepted strategy to address the problem of multiple testing is to adjust the significance level according to the number of tests performed, i.e. Bonferroni's correction. The Bonferroni corrected significance level (α^*) is calculated by;

$$\alpha^* = \alpha/n$$

where α denotes the traditionally desired significance level (e.g. 0.05) and n the number of tests performed. Adjustment according to Bonferroni was done in all three papers. It can be debated whether this method is too conservative as many of the markers tested are not independent of each other due to strong LD, and the possibility of falsely accepting the null hypothesis (type II error) is present. In paper II only correction for markers tested at each locus (e.g. *HLA-B*), and not across all loci, was done to apply a less strict correction due to the strong LD between several HLA alleles. Previously established HLA associations in PSC were *a priori* hypothesized to be associated with small duct PSC and subgroups with and without IBD in paper II, and uncorrected *P*-values below 0.05 were therefore considered significant for these alleles. However, if small duct PSC and subgroups are considered distinct entities, one can argue in favor of applying Bonferroni's correction. In paper III no correction was performed for the HLA analyses and uncorrected *P*-values were presented. Although a less strict correction, or no correction, increase the risk of false positive findings, it also minimize the risk of type II errors.

Statistical power to detect a genetic association is influenced by several factors in addition to characteristics of the study population as outlined in section 4.1. This includes number of cases and controls included in the analysis (i.e. sample size), control to case ratio, effect size (i.e. OR) of the associated marker, frequency of the marker and the significance level selected to define an association. Figure 12 illustrates the correlation between power and these factors.

An increased sample size, increased control to case ratio, large effect size, high frequency of the tested marker in the control population and a non-strict significance level will all contribute to increased power. A power threshold of 0.80 is typically used in genetic association studies, insufficient power predispose to type II errors. Power calculations were done pre-analytic as well as post-analytic in paper II to assess number of individuals needed for inclusion, and to evaluate results in subgroup analyses, respectively. In all three studies included in this thesis, sample size represents the biggest obstacle to achieve sufficient power, especially in paper II and III.

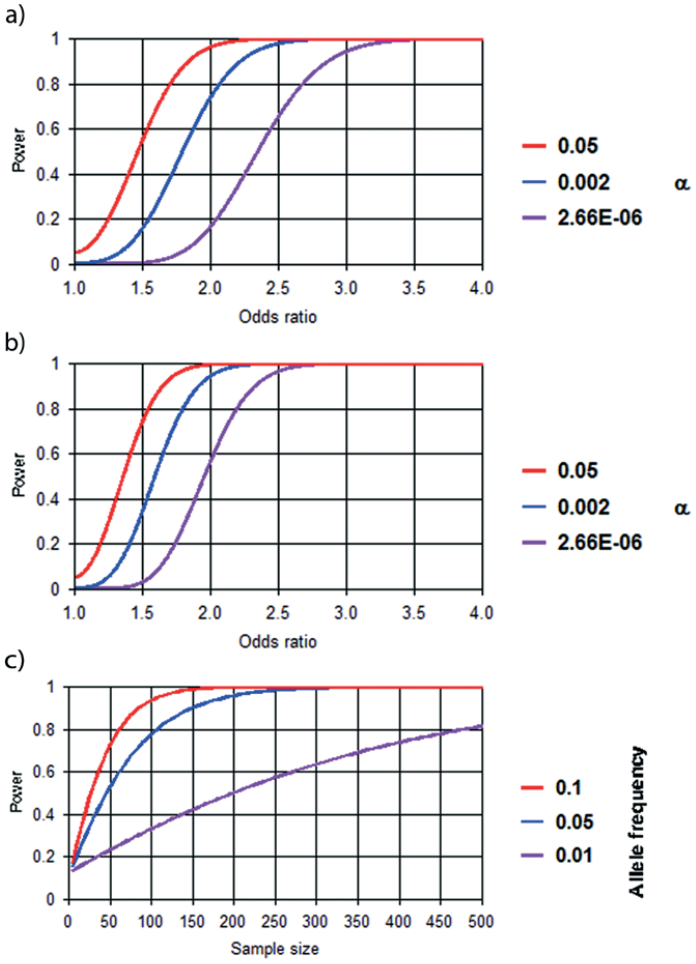

Figure 12: The correlation between power, odds ratio, sample size, allele frequency and control-to-case ratio.

Panel a highlights the variation in power for different odds ratios (ORs) with significance level (α) of 0.05, 0.002 and 2.66 x 10^{-6} (a study specific significance level of 2.66 x 10^{-6} was set in the univariate logistic regression analyses in paper I according to the number of comparisons). The allele frequency was set to 0.10, number of cases to 500 and a control-to-case ratio of 1 was chosen. Panel b is identical to panel a except the control-to-case ratio which was set to 3. Panel c shows how power varies with sample size and allele frequencies of 0.10, 0.05 and 0.01, respectively. The significance level was set to 0.05, the OR to 3, and the control-to-case ratio to 3. These power estimates assume usage of a chi-squared test. Power estimates and graphs were done in PS: Power and Sample Size Calculation, version 3.0.43 (143).

4.7.3 Haplotype estimation and LD measures

The two terms haplotype and LD are introduced in section 1.3.1. Information on parental haplotypes is necessary to accurately deduce the haplotypes of an individual. However, in a case-control study design, where no family data exists, it is possible to estimate haplotype frequencies from genotype frequencies by various algorithms. Estimation of SNP haplotypes was done prior to imputation in paper I by MaCH v.1.0 (136). Estimation of *HLA-B-DRB1* haplotypes in paper II was done in PHASE v.2.1 (144, 145). Calculation of LD between genetic markers (e.g. between SNPs, HLA loci or SNPs and HLA loci) can help to determine the extension of a haplotype and is important when interpreting an association to a genetic marker. This is exemplified in paper I where LD calculations showed strong LD between the associated HLA-B*08 allele and alleles of other loci on the same haplotype, located far apart. This could indicate that the association found at *HLA-B* may simply tag other alleles on the same haplotype, and thus not represent the true causative variant. LD calculations were done in PLINK (paper I) and in UNPHASED (paper II).

4.8 Ethical aspects

Written informed consent was obtained from all study participants still alive. Exemption from consent was granted by the Regional Committee for Research Ethics in Southern Norway for all liver recipients deceased at time of study initiation (paper III). Furthermore, The Norwegian Health Directorate approved the utilization of DNA samples from deceased liver donors (paper III) and PSC patients (paper I and II). Ethical approval was obtained by The Regional Committee for Research Ethics in Southern Norway, and from research ethics committees at each collaborating center abroad, in accordance with the declaration of Helsinki.

5 GENERAL DISCUSSION

This thesis largely concerns refinement of the genetic susceptibility within the MHC to PSC and the clinical subphenotype of small duct PSC, with an emphasis on the classical HLA class I and II genes (paper I and II). In addition, the relationship between classical HLA genes and risk of acute rejection of the transplanted graft was studied in patients treated with liver transplantation.

5.1 Genetic contribution to PSC risk and potential functional relevance

The genetic risk pool identified in PSC so far (listed in Table 6) consists of 16 genome-wide significant loci (including the MHC). They are collectively estimated to account for 7.3 % of the overall PSC risk (45) and 12.1 % of the estimated genetic heritability of PSC (146), reflecting that a huge fraction (~85 %) of the genetic contribution to PSC risk is unidentified so far. The superior role of MHC over other genetic loci is evident, underscored by their different effect sizes (OR = 2.83 *vs.* ORs between 1.15-1.39, respectively) (41, 43, 45), however the proportion of the genetic risk explained by MHC in PSC is unknown. In other diseases like coeliac disease and type 1 diabetes, the MHC association has been estimated to constitute ~50 % of the total genetic risk (90, 147).

Table 6: 16 genome wide significant genetic loci associated with primary sclerosing cholangitis.
Genes in bold represent genes involved in T cell regulation. Adapted from (51).

Locus	Suggested candidate gene(s)
1p36	*MMEL1*, ***TNFRSF14***
2q13	*BCL2L11*
2q33	***CD28***
2q37	*GPR35*
3p21	*MST1*
4q27	***IL2, IL21***
6q15	*BACH2*
6p21	**Classical HLA class I and II genes**
10p15	***IL21RA***
11q23	***SIK2***
12q13	***HDAC7***
12q24	*SH2B3*
18q21	***TCF4***
18q22	***CD226***
19q13	***PRKD2***
21q22	*PSMG1*

It is important to remember that none of the 16 associated loci are exclusively necessary to cause disease (indicated by a 100 % frequency in patients). Based on results from paper I, as many as 93 % of the PSC patients carry one, or several of the four risk HLA haplotype

markers outlined in Figure 17 (i.e. HLA-B*08, HLA-DRB1*13:01-DQB1*06:02, HLA-B*07, HLA-DR2-DQB1*06:02). However, the corresponding frequency in healthy controls was 66 %, proving that none of these HLA risk haplotypes are disease specific, and that other additional factors must be present to initiate disease. Risk variants at the 15 non-MHC loci are also to a large extent present in healthy controls with risk allele frequencies ranging from 10-84 % (41, 43, 45). Taken together, identified genetic risk variants in PSC are only contributing pieces to a large pathogenic puzzle.

Although the 15 non-MHC loci are far less complex than the MHC, it is still a challenge to pinpoint causal variants for many of these loci, as they are multigenic with considerable degree of LD. Several strategies may be applied in order to identify the causal variant for genetic associations, including fine-mapping strategies (as performed in paper I), methods to postulate relevance of the associated variant at protein level and prior knowledge of gene functions at the locus in question based on published literature. However, none of these approaches prove causality, and merely represent hypotheses. Functional studies must be undertaken to establish a role in disease pathogenesis.

Bearing in mind the limitations described above, especially prior published literature, proposed candidate genes at the various 15 non-MHC loci (51) and in the MHC are listed in Table 6 and the possible site of action for the candidate genes is illustrated in Figure 13. The majority of non-MHC loci encode proteins who directly or indirectly affect T cell biology.

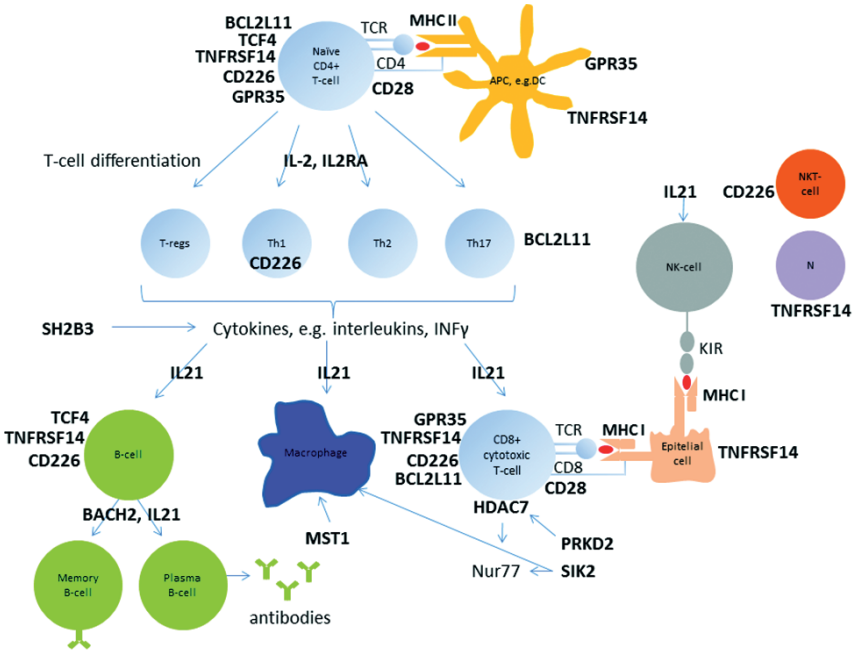

Figure 13: Site of action for suggested candidate genes in primary sclerosing cholangitis.
Possible candidate genes are marked in bold. Most genes are directly or indirectly involved in T cell (marked in blue) biology. However, cells of innate immunity are also involved (i.e. antigen presenting cells (APCs) such as dendritic cells (DC) and macrophages, neutrophils (N), natural killer (NK) cells and natural killer T cells (NKT cells).

In the MHC classical HLA genes are strong candidate genes in PSC (supported by results in paper I and discussed in section 5.2), as it is for most MHC associated conditions. The most prominent hypothesized molecular mechanism for genetic association of classical HLA genes is their peptide presenting ability to T cells. The polymorphism of the genes encoding HLA molecules is mainly confined to the peptide binding groove, enabling different set of alleles to bind and present different range of peptides to T cells. Disease associated HLA molecules may exert their function both in the thymus where they mature, and in the periphery at the site of organ damage. In order to reveal the molecular mechanisms behind the peptide-presenting hypothesis, it is crucial to identify and access disease triggering/driving antigens as well as the T cells involved in the immune response at the site of tissue damage. However, in most conditions, including PSC, the disease relevant antigen(s) is unknown and the T cells are hard, or even impossible, to access due to anatomical localization of the affected organ. Yet, there

are conditions in which these criteria are met, which support the peptide-presenting hypothesis. The most notable example being coeliac disease, an inflammatory disease of the small intestine, in which a primary HLA locus (*HLA-DQ*) has been identified (90), a disease relevant antigen is known (i.e. gluten) and reactive CD4+ T cells towards modified gluten peptides in the gut mucosa of patients exist (148). Although peptide presentation to T cells is the most plausible molecular mechanism by which HLA molecules are involved in disease pathogenesis, other possible mechanisms exist. HLA molecules are able to interact with receptors in a peptide independent manner, acting as ligands rather than peptide presenting molecules. The most prominent example being interaction with HLA class I molecules and KIRs. This interaction influence the activation status of the NK cell, and genetic studies have indicated that this interaction is of importance in viral infections, autoimmune- and immune mediated disorders, cancers, reproduction and organ transplantation (149). Genetic association of HLA class I genes as KIR ligands has been reported for PSC (80, 106), and a potential role of this interaction could not be excluded in paper I, although it did not seem to be of superior importance. Aberrant processing of the HLA-peptide complex and different surface expression of particular HLA alleles may be other molecular mechanisms behind genetic association with classical HLA genes. Possible molecular mechanisms of genetic association with classical HLA genes are schematically illustrated in Figure 14.

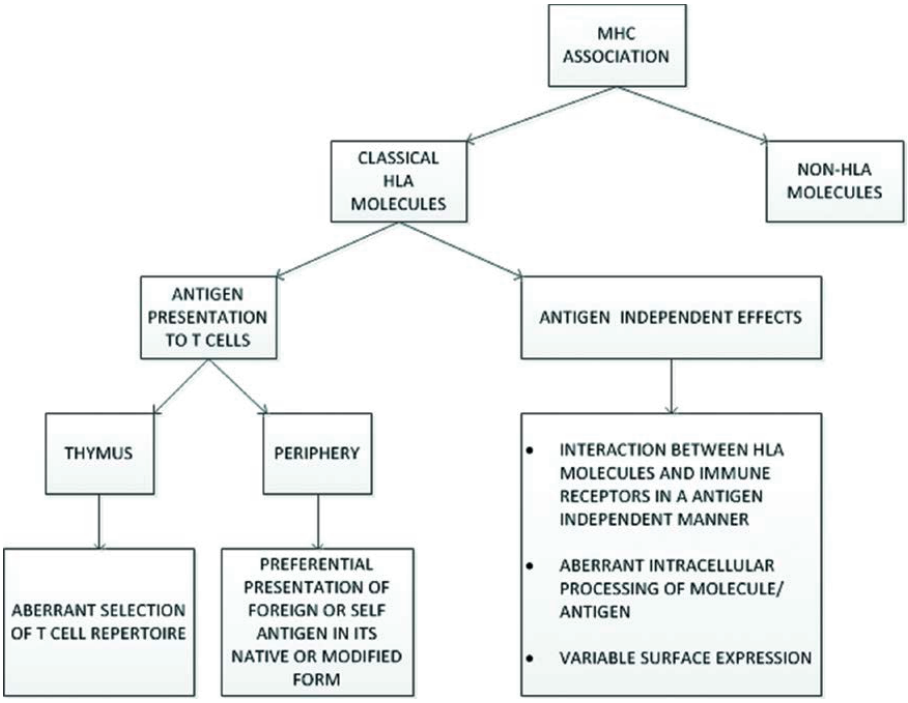

Figure 14: Overview of potential molecular mechanisms underlying genetic associations with the major histocompatibility complex.

The likely major role of classical HLA genes and their functional role as peptide presenting molecules to T cells, together with proposed functions of the non-MHC loci, suggest a major involvement of adaptive immune responses (CD4+ and CD8+ T cells, B cells and antibodies) in PSC pathogenesis. These genetic results are supported by the dominant infiltration of T cells (both CD4+ and CD8+) in PSC livers (150, 151). However, it must be emphasized that these biopsies are often taken in patients with advanced or end stage disease and thus potentially mirror a different pathogenic picture than the one at disease initiation.

Most of the 16 associated PSC loci are also associated with other immune mediated diseases. Based on this shared genetic risk across phenotypes, called pleiotropy, additional 33 *suggestive* genetic loci have been identified in PSC (45), many of which have functions that can be related to PSC pathogenesis. It is however important to remember that many of the associated genetic loci in PSC have been identified through studies specifically searching for immunological effects (e.g. the Immunochip study design (45) and candidate gene studies of

HLA genes), perhaps over-estimating the overall role of immunological *vs.* non-immunological effects to PSC genetic architecture.

5.2 Mapping the MHC association in PSC

5.2.1 Challenges in dissecting the MHC association signal

Mapping efforts of MHC association signals have been done in numerous MHC associated conditions. In PSC the complexity of the MCH association signal is substantial, as demonstrated by the MHC SNP association plot in the 2nd PSC GWAS (Figure 15) (41).

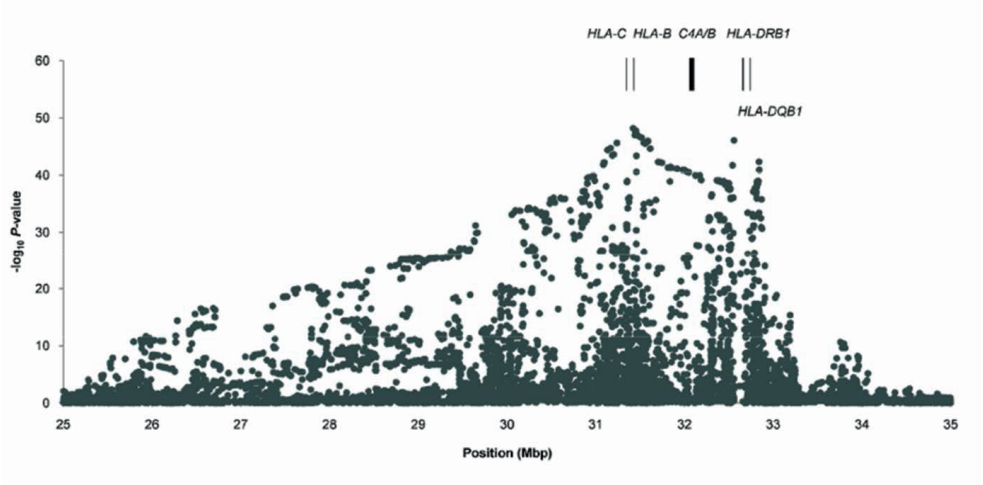

Figure 15: Primary single nucleotide polymorphism association plot within the major histocompatibility complex from the 2nd primary sclerosing cholangitis genome-wide association study.
The strength of the association signal, $-\log_{10}P$-value (y-axis), is plotted against the position on chromosome 6 shown in million base pairs (x-axis). Positions of key human leukocyte antigen (HLA) class I and II genes, i.e. *HLA-C, B, DRB1* and *DQB1* are indicated, in addition to the complement gene cluster (*C4A/B*) in the class III region. Adapted by permission from Macmillan Publishers Ltd: Nature Genetics (41), copyright (2011).

The primary association plot from paper I demonstrated the same complexity, and numerous of variants, including SNPs, HLA alleles and amino acid variants of DRβ, reached study specific significance threshold (Figure 16). This high number of significant variants may be confusing and difficult for the researcher to untangle. However, many of these variants are not independent of each other, as they are tightly linked through LD and tag association with the same MHC haplotype.

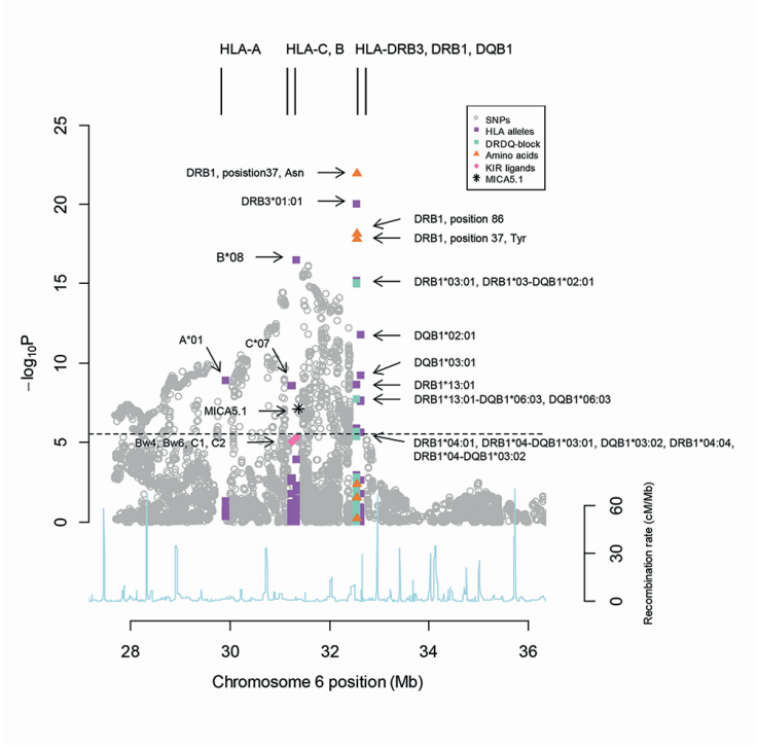

Figure 16: Primary association plot of genetic variants within the major histocompatibility complex in primary sclerosing cholangitis (paper I).

The strength of the association signal, -$\log_{10}P$-value (y-axis), is plotted against the position on chromosome 6 shown in million base pairs (Mb, x-axis) for each of the 18,771 variables tested for association. The black, horizontal, dotted line represents the Bonferroni-corrected statistical significance threshold of $P \leq 2.66 \times 10^{-6}$. Association results are categorized according to single nucleotide polymorphisms (SNPs), classical human leukocyte antigen (HLA) alleles, *DRB1-DQB1* haplotypes, and previously reported associations of amino acids of DRβ, the MICA 5.1 allele and *HLA-B* and *HLA-C* alleles according to their killer immunoglobulin-like receptors (KIRs) ligand ability (i.e. Bw4/Bw6 and C1/C2). Positions are given according National Center for Biotechnology Information's build 37 (hg19). Adapted from Naess *et al.* PLoS One 2014 (152) under the Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/)

Following the primary association analysis, the typical approach in mapping efforts is to select the most significant variant (in our case Asparagine at position 37 of DRβ), and do subsequent conditional analysis, controlling for this variant. This approach is repeated until no more significant associations are left. Such a stepwise conditional approach was also performed in paper I. However, mapping association signals within the MHC is not straight-

forward, and could benefit from several analytical approaches. We therefore chose to explore two additional conditional approaches, i.e. in addition to Asparagine at position 37 of DRβ, HLA-B*08 and DRB3*01:01 were also selected as primary conditional variables based on *a priori* knowledge of haplotype associations in PSC. These three strategies yielded both overlapping and differentiating results, emphasizing that the analytical approach to a large extent affect the outcome. Rather than classifying one of our three approaches as the "correct" one, these were complementary, underlying the importance of exploring several different conditional strategies and their interpretation in the setting of known MHC haplotype architecture and previous HLA association studies of the disease in question. However, potential selection bias is possible if such prior knowledge is accounted for in the analytical approach. It also remains a possibility that other, unexplored, analytical strategies of our data would have given an even more "correct" picture of the MHC association in PSC and possible important effects could have been missed by our chosen strategies. Adding to the discussion, is the need for even more refined statistical analyses of the MHC, as exemplified by our unconditional multivariate logistic regression analyses of the classical HLA genes and alleles, which added valuable, and in part novel, results.

An important aspect when analytical approaches are selected and results interpreted is awareness of the different biological implications of the various genetic variant classes included in the analysis, and their strengths and weaknesses in an analytic setting. Most recent mapping efforts of the MHC typically include three variant classes; SNP genotypes across the MHC, and HLA genes at the level of classical HLA alleles and amino acid positions. Whereas an HLA allele represents the full HLA molecule, an amino acid variant or SNP located within an HLA gene only reflects a small part of the HLA molecule. Greater amount of information is thus captured by allelic information. However, a potential strength of amino acid variants and SNPs is their ability to more precisely define which part of the molecule that is affected, and thus better specify the causal effect(s) within the HLA molecule. An additional strength of SNP genotypes included in MHC mapping efforts is their ability to point to regulatory functions of proteins rather than variation of the protein itself, as many of the SNPs typically are located in non-coding regions. Because disease causing variants once arose due to mutations on ancient haplotypes, an important step in mapping MHC associations is to determine haplotype architecture. The greater amount of information included, the better the associated haplotype is characterized. Thus, multiallelic variants (e.g. HLA alleles) that often consist of hundreds of SNPs, are better suited to capture haplotype structure than one single

biallelic SNP. Lastly, the statistical power to detect associations is greater for biallelic loci than for multiallelic loci due to the number of statistical tests performed. In our conditional mapping effort, all genetic variants were included as dosage data (i.e. 0, 1 or 2 copies), resulting in equal statistical strength to detect association for each variable included. In sum, given their different strengths and weaknesses in an analytical setting, the three different variant classes typically included in MHC mapping efforts are complementary.

Coverage of genetic variation in the MHC in paper I was done by SNP genotypes and HLA genes at the level of classical HLA alleles. A mapping based on amino acid variation of classical HLA molecules was not feasible in paper I, and it is possible that such an effort would have added valuable information to the over-all picture of the MHC architecture in PSC. Genotyped and imputed SNPs in our analyses were confined to the regions outside the classical HLA genes, and it is possible that newer SNP screens with more dense coverage of the entire MHC, including classical HLA genes, would have added additional information to our mapping effort. The most important HLA class I genes, *HLA-A, B* and *C* were covered by inclusion of sequencing data for these loci. However, at MHC class II, only *HLA-DRB3, DRB1* and *DQB1* were included, and thus information on several MHC class II loci was not included (see Table S3 in paper I for annotated loci in MHC class II and coverage in present study). Insufficient coverage of the MHC class II region is also a bias of other MHC mapping efforts, and must be kept in mind when results are interpreted. Taken together, the better the coverage of the genetic variation present, the less likely it is that true associations are over-looked.

5.2.2 *HLA-B* and *HLA-DRB1* – primary risk loci in PSC?

PSC is, as many other conditions, strongly associated with genetic variants within the MHC, but what does this mean? For simplicity this can be narrowed down to association with classical HLA genes *vs.* non-HLA genes. A primary role for the classical HLA genes in most MHC associated conditions is supported by several observations. First, association signals in MHC mapping studies typically peak in the vicinity of classical HLA class I and/or II genes. Second, for many conditions, association with the same HLA allele(s) is observed in populations of different ancestry and thus LD patterns. For example, association with HLA-B27 in ankylosing spondylitis is observed in all populations studied (153). Third, if different HLA alleles are associated in different populations, these may share peptide binding motifs, indicating similar properties of the encoded molecule. For example, in rheumatoid arthritis, several associated *HLA-DRB1* alleles encode a similar stretch of amino acids at position 67-

74, influencing the peptide binding groove of the HLA-DR molecule, referred to as "the shared epitope hypothesis" (96). In addition, work in spontaneous and transgenic mice models in several MHC associated conditions, support a role for classical HLA molecules in disease pathogenesis (154). This may be exemplified by development of ankylosing spondylitis like disease in rats made transgenic for HLA-B27 (155) or development of spontaneous diabetes in transgenic mice that express the diabetes associated human HLA-DQ8 (i.e. HLA-DQA1-03:01-DQB1*03:01) together with the co-stimulatory molecule B7-1 (CD80) (156).

A primary role for the classical HLA genes in PSC was suggested in paper I. This is supported by previously published MHC SNP association plots in PSC (40, 41, 45), which peak in the vicinity of classical HLA genes. By combining two different analytic strategies, in total nine MHC haplotypes were associated with either risk (n = 5) or protection (n = 4) against PSC. On eight of these haplotypes, the strongest associated variant was located at either HLA class I or HLA class II, represented by either a classical HLA allele or amino acid variation of DRβ. However, due to the complex and variable LD pattern of haplotypes harboring these associated HLA class I and II variants, it is important to explore their extension in order to adequately interpret the results (illustrated in Figure 17). As anticipated, the various haplotypes showed variable extension, with haplotype 1 (representing AH8.1) being the most extended. One cannot rule out the possibility that the haplotype markers are indirectly associated and only tag true causative variants. However, it is likely that true causative variants are in strong LD (r^2>0.8) with these markers, represented by the extension of the black colored line in Figure 17. With that in mind, it can be argued that for eight of the haplotypes true causative variants are likely to be located in, or in close vicinity, of either MHC class I or II. Based on results from paper I, alleles at *HLA-B* or *HLA-DRB1-DQB1* represented the strongest associated variants on these eight haplotypes, and favor these as primary risk loci. Due to strong LD between *HLA-DRB1* and *HLA-DQB1*, it is difficult to predict which of these neighboring loci that represent the most likely causal variant. However, for two risk haplotypes, the *DRB1* allele (DRB1*13:01 and DRB1*15:01) showed superior association relative to both the *DRB1-DQB1* haplotype (DRB1*13:01-DQB1*06:03 and DRB1*15:01-DQB1*06:01) and the *DQB1* allele (DQB1*06:03 and DQB1*06:02) in the conditional analytic approach, thus suggesting a primary role for the *DRB1* locus. This is supported by the association with DRB1*13:01 in PSC patients of African decent (110), a population in which DQB1*05:01 and DQB1*06:03 are observed on the DRB1*13:01

haplotype at comparable frequencies (157). For other haplotypes, such as the protective DRB1*04-DQB1*03, a superior role at *DRB1* is not as evident.

Although a primary role for classical HLA genes was indicated in paper I, previous and current results suggest additional contribution of non-HLA genes to the MHC association in PSC. A novel and independent SNP association in the class III region, centromeric of neurogenic locus notch homolog 4 (*NOTCH4*), was found in paper I. To establish this as a true association, replication is warranted. The potential functional relevance of this finding can only be speculated, however the diverse biological function of the notch signaling pathway as well as association with this region in other MHC associated conditions, such as schizophrenia (158) and type 1 diabetes (159) is interesting. Previous reported association with the MICA 5.1 allele (87, 106, 107) located centromeric of *HLA-B*, was explored in paper I. Although this variant reached study specific significance threshold in the primary association analysis (Figure 16), it did not seem to be of primary importance and previous association with the MICA 5.1 allele could be attributed to its presence on the PSC associated haplotypes; AH8.1 and AH7.1. Yet, weak, additional contribution from this locus cannot be excluded. As for the MICA 5.1 allele, a primary role for the previous reported HLA-C and HLA-Bw KIR ligands (80, 106) was not found in paper I. However, one cannot rule out contribution from these variants as well, especially a potential epistatic effect between HLA class I molecules and KIRs, which was beyond the scope of this study.

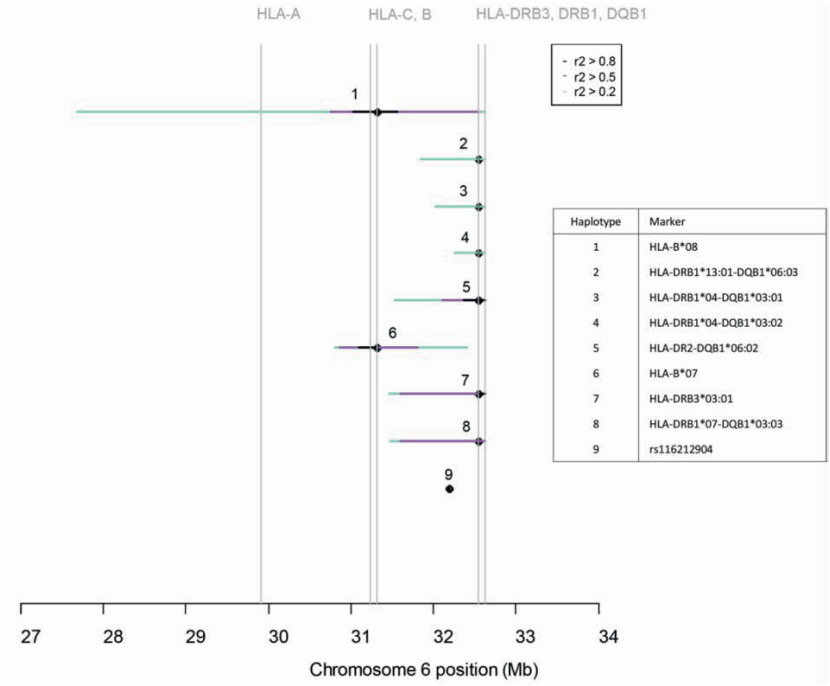

Figure 17: The extension of nine significantly primary sclerosing cholangitis associated major histocompatibility complex haplotypes (paper I).

The extension of the haplotypes is determined by linkage disequilibrium (LD) between a single haplotype marker (listed in figure box) and any single nucleotide polymorphism (SNP) in the dataset. Strong (r^2>0.8), intermediate (r^2>0.5), and moderate (r^2>0.2) LD is indicated by black, purple and green lines, respectively. Haplotype 1; AH8.1 (A*01-C*07-B*08-DRB3*01:01-DRB1*03:01-DQB1*02:01). Haplotype 2; HLA-DRB1*13:01-DQB1*06:03. Haplotype 3; HLA-DRB1*04-DQB1*03:01. Haplotype 4; HLA-DRB1*04-DQB1*03:02. Haplotype 5; HLA-DRB1*15:01-DQB1*06:02. Haplotype 6; HLA-B*07. Haplotype 7; HLA-DRB3*03:01-DRB1*13:02-DQB1*06:04. Haplotype 8; HLA-DRB1*07:01-DRB1*03:03. Haplotype 9; an independent LD-block centromeric of *NOTCH4*, tagged by rs116212904. DRB1*15:01 is part of the DR2 serotype group. Haplotype 1, 2, 5, 6 and 9 represent risk haplotypes, whereas 3, 4, 7 and 8 represent protective haplotypes. Adapted from Naess *et al.* PLoS One 2014 (152) under the Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/). In the figure box the haplotype marker for haplotype 4 is corrected to HLA-DRB1*04-DQB1*03:02 from HLA-DQB1*04-DQB1*03:02 (as originally presented in the paper).

5.2.3 The ancestral haplotype 8.1

In PSC the complexity of the primary association signal is explained by association with several MHC haplotypes of variable extension, most notably by variants tagging the AH8.1.

This can be visualized in Figure 16, as HLA-A*01, C*07, B*08, MICA 5.1, DRB3*01:01, DRB1*03:01, DQB1*02:01, Asparagine at position 37 of DRβ and the top SNP (rs139345387), all tag, with variable efficiency, this haplotype. Ancestral haplotypes are highly conserved haplotypes which are believed to origin from a common remote ancestor. The AH8.1 is one of the longest haplotypes in the human genome at 4.7 million nucleotides in length, starting at ~ 1 million nucleotides telomeric of *HLA-A*, expanding ~800 000 nucleotides centromeric of *HLA-DQB1*, with in total 311 annotated loci (160). Classical HLA alleles found on AH8.1 are HLA-A*01:01, C*07:01, B*08:01, DRB1*03:01, DQA1*05:01 and DQB1*02:01. The occurrence of AH8.1 is high among people of Caucasian decent, with a frequency of ~ 10 % (http://www.allelefrequencies.net/). The high frequency and strong conservation indicate a fairly recent historic evolutionary advantage of AH8.1, perhaps increased resistance to infectious diseases. Yet, this "hyper-responsive" collection of genes may at the same time predispose to autoimmunity.

The AH8.1 is associated with a large number of immunopathological conditions. With > 300 genes, of which a third have immunological function, it is reasonable that a combined effect of multiple genes and their potential interaction(s) influence the immune response conducted by this haplotype. Several mechanisms have been proposed as contributing factors besides the antigen-specific mechanisms, including a shift towards type 2 cytokine production, which favors antibody responses (161). There is no common region of this haplotype that is of primary importance in all AH8.1 associated conditions. For some diseases the primary contribution of this haplotype is confined to the class II region, as for instance with *HLA-DQA1-DQB1* in coeliac disease and with the *HLA-DRB1-DQB1* region in type 1 diabetes. For other conditions, such as PSC, associations have been found for gene alleles expanding almost the entire haplotype, making it difficult to identify primary causal genes.

Our analyses from paper I support a primary contribution for HLA-B*08:01 on AH8.1 in PSC. The multivariate unconditional regression which included analysis of the *HLA-A, C, B, DRB3, DRB1* and *DQB1* alleles, resulted in significant association with only the HLA-B*08 allele. Furthermore, HLA-B*08 associated second most strongly out of the AH8.1 HLA alleles, after the DRB3*01:01 allele, in the primary association plot from the univariate conditional strategy (Figure 16). The superior association of DRB3*01:01 from this analysis is likely attributable to its presence on another PSC risk haplotype; DRB1*13:01-DQB1*06:03. Selecting HLA-B*08 as the first conditional variable clearly reduced the complexity of the MHC association signal, leaving no significant associations with variants of the AH8.1 and a

residual signal shifted towards the class II region. Our results favoring HLA-B*08 as the primary locus on AH8.1 in PSC, is supported by several other observations. First, the MHC SNP peak was located in/or in the vicinity of *HLA-B* in both genome-wide association studies in PSC (40, 41), as well as in the more recently published Immunochip study (45). Furthermore, HLA-B*08, and not DRB1*03:01, associated with PSC in a study population of African American origin (110), which represent and older population than Scandinavians, with lesser degree of LD. However, although *HLA-B* is most likely the primary locus on this haplotype in PSC, we cannot exclude the possibility that a variant in strong LD with *HLA-B* is the actual causal variant. It is also reasonable to assume that other variants on AH8.1, including both HLA and non-HLA genes contribute to the increased disease risk in PSC.

5.3 The influence of MHC associations on clinical subphenotypes

The clinical picture of PSC (i.e. phenotype) is diverse, underscored by variable presence of co-morbidities (such as presence of IBD and other immune-mediated diseases, overlap with AIH, and cancer development), autoantibodies (e.g. anti-neutrophil cytoplasmatic antibody (ANCA)), IgG4 antibody, different localization of the bile duct damage (small duct PSC *vs.* classic/large duct PSC) and a variable rate of disease progression. This clinical diversity among PSC patients is likely reflecting, inter-individual variation in two categories; genetic susceptibility and environmental exposure. Given the superior role of the MHC association in the overall genetic predisposition to PSC, it is reasonable to argue that inter-individual variations in MHC susceptibility play a major role in the first category, thereby contributing to phenotypic diversity. To explore whether different MHC susceptibility could be linked to the phenotypic subgroup of small duct PSC, alleles at four key HLA loci (i.e. *HLA-A, B, C* and *DRB1*) were investigated in paper II.

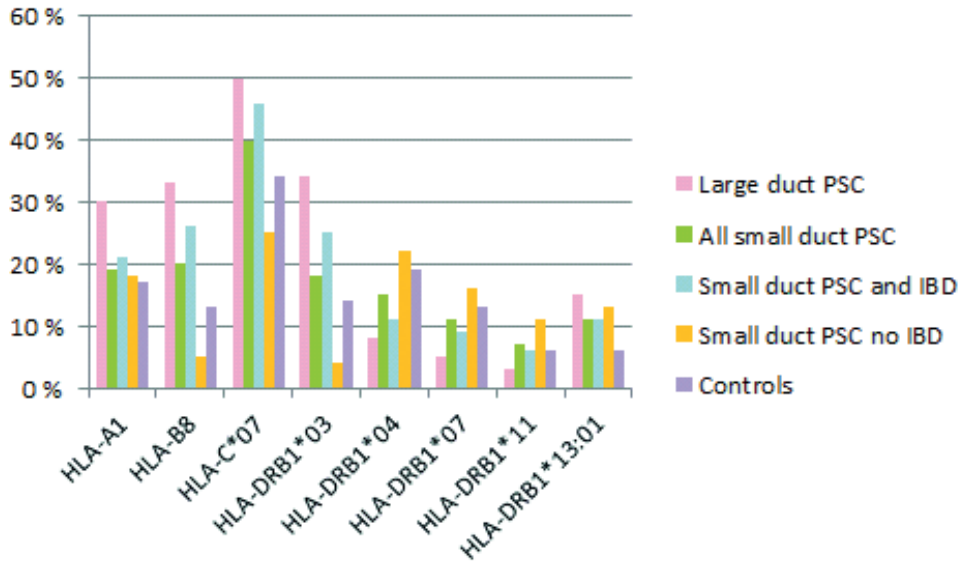

Figure 18: Human leukocyte antigen allele frequencies from paper II.

Frequencies of eight human leukocyte antigen (HLA) alleles associated with large duct primary sclerosing cholangitis (PSC) compared with all small duct PSC, small duct PSC with co-existing inflammatory bowel disease (IBD), small duct PSC without co-existing IBD, and healthy controls.

Figure 18 demonstrates frequencies of the eight HLA alleles significantly associated with large duct PSC in paper II, all of which have previously been reported. Only two of these, HLA-B*08 (serotype B8) and HLA-DRB1*13:01, were significantly associated (both conferring risk) in the overall small duct PSC patient population. However, stratification of the small duct PSC group based on co-existing IBD, showed that the subgroup with co-existing IBD greatly resembled large duct PSC in its HLA association whereas the group with no co-existing IBD only associated with one HLA allele; DRB1*13:01. These results may suggest that small duct PSC patients without IBD represent a distinct group with differentiating pathogenesis. Due to the rarity of small duct PSC (~10 % of the entire PSC population), an obvious limitation of our study is the small number of small duct PSC patients included (n = 87), and even smaller number of patients in subgroup analyses. Our results should ideally be replicated in a larger population cohort, and preferentially also in a study population of different ethnicity. We speculated, based on our genetic data, whether progression to large duct affection (which occurs in ~25 % of the patients initially diagnosed

with small duct PSC) is confined to small duct patients with co-existing IBD. To confirm this follow-up data must be registered and included, which was beyond the scope of our study.

Other PSC clinical subphenotypes than small duct PSC, have recently shown to have a different HLA predisposition than the PSC group as a whole. Elevated serum IgG4 level is seen in a proportion of PSC patients (~9-15 %), and has previously been associated with a more severe disease course (162). Recently, it was shown that high IgG4 levels are associated with reduced frequency of HLA-B*08 and increased frequency of the DR2-haplotype (HLA-B*07-DRB1*15:01) when compared to patients with low levels of IgG4 (163). Furthermore, ANCA positive PSC patients, (reported in 43-93 % of the PSC population, reviewed in (164)) associate more strongly with the AH8.1 alleles; HLA-B*08 and DRB1*03:01, than ANCA negative PSC patients (Hov *et al.*, unpublished). ANCA positive PSC patients were also shown to have an earlier disease onset and a reduced frequency of biliary cancer in the same study. Our results from paper II indicated a different HLA predisposition in small duct PSC with and without co-existing IBD. In contrast, no such difference was reported in a study comparing *HLA-DQB1* and *DRB1* alleles in large duct PSC patients with and without IBD (81). Yet, there was a trend towards reduced frequency of DRB1*04 and an increased frequency of DRB1*15:01 in PSC patients with co-existing IBD. Supporting this, is our results from paper II, where a significant reduced frequency of DRB1*04 in large duct PSC with IBD *vs.* large duct PSC with no IBD, was found (OR = 0.41, 95 % CI: 0.24-0.68, P = 0.0007). This could indicate a potential protective effect of DRB1*04 *vs.* IBD status in large duct PSC. Earlier studies have also linked rapid disease progression in PSC with particular HLA alleles, i.e. DRB1*04, DRB3*01:01 (79, 165), yet others failed to replicate this finding (166). In addition, the DRB1*03:01-DQB1*02:01 heterozygote genotype has been reported to associate with an accelerated progression of PSC (167).

Clinical subphenotypes of other MHC associated diseases have also shown variation in HLA susceptibility, in line with observations in PSC. Association with *HLA-DRB1* alleles in anti-citrullinated peptide antibodies (ACPA) positive, and not in ACPA negative rheumatoid arthritis patients (168), is an example of this. Different HLA association according to early or late onset of myasthenia gravis (100), and a variable risk gradient for HLA-DRB1*07:01 and DRB1*01:03 in Crohn's disease according to localization of disease (i.e. ileal and/or colonic involvement) (93), are other examples. Taken together these data indicate that the diverse clinical picture of PSC (and other conditions), at least partly, can be linked to variable genetic predisposition within the MHC.

5.4 Do shared HLA associations indicate etiopathological similarities between PSC and acute rejection of liver grafts?

Rejection of an allograft can be defined as "an immunologic reaction to the presence of a foreign tissue or organ that has the potential to result in graft dysfunction or failure" (169). Since the beginning of solid organ transplantation, it has been known that liver grafts are better tolerated by the recipient's immune system and less prone to rejection compared to other organs (170). The mechanism(s) for this immunological advantage of the liver is unclear (170). The most common type of rejection seen after liver transplantation is acute rejection (170), reported to occur in 20-40 % (25). Although acute rejection is quite easily reversed in liver recipients, it still represents a clinical problem as these patients are in need of additional immunosuppressive therapy with potential harmful side effects.

Presence of acute rejection seems to be correlated to the underlying liver condition, and has been reported in excess in patients transplanted on the basis of immune mediated liver conditions, such as PBC, AIH and PSC. Conversely, patients with non-immunological conditions, such as alcoholic liver disease, seem to be protected against acute rejection (71, 171, 172). Results from paper III were in line with these previous reports, with an increased frequency of acute rejection in PSC patients (41%) *vs.* non-PSC patients (23%). These observations reason well with the presence of a more active immune system in patients with immune mediated diseases and the possibility that immunological mechanisms responsible for the primary liver disease itself may also contribute to rejection mechanisms after liver transplantation. Given the strong HLA association in immune mediated liver diseases, one such mechanism might be the interaction between HLA molecules and T cells, the latter being prominent in cellular infiltrates of acute rejection (115). The link between particular HLA alleles in the recipient and risk of acute rejection, as shown in paper III, supports this assumption. HLA-B*08 and HLA-C*07 (both located on AH8.1), associated with risk and HLA-DRB1*04 with protection of acute rejection, even when PSC patients, with known association with these alleles, were removed from the analysis. Hypothetically this association could be driven by patients with AIH, which have a reported association with alleles of AH8.1 (173). However only 2 (out of 7) individuals with AIH carried alleles of AH8.1 in the non-PSC group. The small size of the groups with acute rejection, both in the overall recipient population ($n = 41$) and in the non-PSC group ($n = 23$), and uncorrected P-values reported for the HLA alleles represent limitations to our study. It remains a possibility that the reported associations with HLA alleles and risk of acute rejection represent false positive findings and

replication of our results is necessary. However, a link between alleles of AH8.1 and risk of acute rejection in liver recipients has previously been reported (174, 175), and support our results. The shared association of AH8.1 in acute rejection and autoimmune diseases, could indicate etiopathological similarities between immune responses towards the allograft (i.e. alloreactivity) and self-tissue (i.e. autoimmunity).

Several studies have reported an association between risk of acute rejection after liver transplantation and increasing number of HLA mismatches between donor and recipient (172, 175-177). In line with previous results (178), we could not replicate these findings, as no association between HLA matching and acute rejection, either within each locus (*HLA-A, B, C* and *DRB1*), or collectively, was found in paper III. As with acute rejection, conflicting results exist as to whether HLA matching is beneficial or not with respect to patient and graft survival, and various studies have reported positive (178), negative (179, 180) and no (181) effect on survival. A possible explanation for the conflicting results could be the suggested opposing effect of HLA matching in autoimmune and non-autoimmune recipients (175, 176). The reported beneficial effect of HLA mismatch on graft survival in autoimmune recipients could be attributed to less autoimmunity towards the transplanted graft as fewer self antigens are present.

Based on our results, the HLA background of the recipient seems to have bigger influence on risk of acute rejection after liver transplantation than HLA matching of recipient and donor. This may favor immunogenetics of the recipient as major determinants of alloreactivity following liver transplantation, and help explain why recipients with immune mediated liver diseases are more prone to acute rejections. Searching for host genetic factors affecting risk of rejection may hold potential for future identification of individuals at risk of rejection and personalized immunosuppressive therapy. Greater knowledge of the pathophysiology causing allorecativity could also gain knowledge of autoimmune mechanisms, as these two processes seem to share many features.

6 CONCLUSIONS ON THE MHC ASSOCIATION IN PSC AND FUTURE STUDIES

The main work of this thesis explores the genetics of the MHC at chromosome 6p21, in particular of classical HLA class I and II genes, in PSC. We were not able to find one single MHC effect in our Caucasian study population (paper I). Rather several MHC haplotypes, both protective and risk, were associated with PSC. Due to the issue of LD, we are not able to pinpoint the causative variants on the associated haplotypes with certainty, yet evidence point to *HLA-B* and *HLA-DRB1* as loci of superior importance. Multiple associated MHC haplotypes in PSC could reflect the phenotypic diversity observed in these patients. Results from paper II, which studied key HLA loci in small duct PSC, together with HLA association studies of other PSC subphenotypes, support this.

There is no doubt that the strongest genetic risks for PSC resides within the MHC. The PSC genetic architecture with the towering association within the MHC, mirrors antigen-restricted autoimmunity seen in for instance coeliac disease (Figure 19). A main focus of future MHC related studies in PSC, given the likely primary role of HLA molecules and their ability to present peptides to T cells, will be to detect PSC causing and/or driving antigens. In doing so, several parallel approaches are needed and multiple efforts are currently on its way, or being initiated.

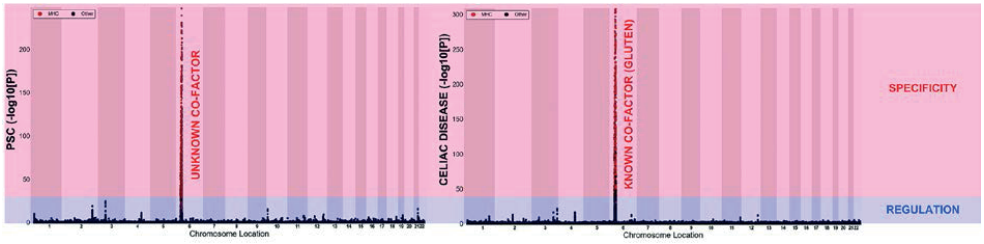

Figure 19: Genome-wide single nucleotide polymorphism association plot illustrating the dominant major histocompatibility complex association at chromosome 6 in primary sclerosing cholangitis (left) and coeliac disease (right).

The strength of the association signal, $-\log_{10}P$-value (y-axis), is plotted against the chromosome location (y-axis). The figure was created and kindly shared by Tom H. Karlsen.

Identification of pathogenic PSC antigen(s) involves work on both sides of the MHC/peptide-receptor interaction. Better refinement of the HLA background can now be done. Recently

established next-generation sequencing based method of the MHC (182) with complete sequencing of HLA class I and II, will provide higher resolution of classical HLA alleles as well as additional information at loci not previously genotyped. Such studies should preferably be conducted in admixed populations to better cope with the issue of LD. Once PSC related HLA alleles have been defined, e.g. HLA-B*08, HLA-DRB1*13:01, their encoded molecules can be isolated from PSC livers of individuals carrying these genetic variants, and subsequent elution of their bound peptides can be done. Another way of identifying candidate antigen(s) is to isolate liver-infiltrating plasma cells from PSC livers and study their produced antibodies. Plasma cells are specific for the antigen that initially triggered activation of their precursor B cell, and so defining the specificity of plasma cell produced antibodies may identify disease relevant antigen(s). Characterization of B- and T cell receptor clonality in PSC livers could also point to existence of disease causing/driving antigen(s), as presence of specific clone(s) point to specific antigen(s) driving the differentiation. Identification of disease specific receptor clones may be used to search for compatible antigens.

Given the reported genetic significance of HLA class I molecules as ligands for killer immunoglobulin-like receptors in PSC (80, 106), further studies on the genetics of the HLA class I–KIR interaction is needed. Advancement in KIR genotyping, also capturing copy number differences, is available (used in paper III) (132). This has led to establishment of KIR imputation algorithms (Vukcevic, D. *et al.*, in submission), making it possible to infer KIR genotyping data for large study populations based on SNP data, parallel to HLA imputation methods. This cost-effective method allows for better characterization of the KIR part of the HLA class I-KIR interaction as well as proper statistical interaction analysis between HLA class I and KIRs, not previously done. If such genetic efforts support the importance of this molecular interaction in PSC, relevant functional studies must be explored.

More than 30 years has passed since classical HLA genes were first associated with PSC (1, 2). Several efforts are now finally being made towards translation of this genetic finding into molecular understanding. Hopefully, these efforts will result in a profound advancement in understanding PSC pathogenesis and eventually lead to new therapeutic opportunities for these patients.

7 REFERENCES

1. Schrumpf E, Fausa O, Forre O*, et al.* HLA antigens and immunoregulatory T cells in ulcerative colitis associated with hepatobiliary disease. Scand J Gastroenterol. 1982;17(2):187-91.
2. Chapman RW, Varghese Z, Gaul R*, et al.* Association of primary sclerosing cholangitis with HLA-B8. Gut. 1983;24(1):38-41.
3. Molodecky NA, Kareemi H, Parab R*, et al.* Incidence of primary sclerosing cholangitis: a systematic review and meta-analysis. Hepatology. 2011;53(5):1590-9.
4. Boonstra K, Beuers U, Ponsioen CY. Epidemiology of primary sclerosing cholangitis and primary biliary cirrhosis: a systematic review. J Hepatol. 2012;56(5):1181-8.
5. Boberg KM, Aadland E, Jahnsen J*, et al.* Incidence and prevalence of primary biliary cirrhosis, primary sclerosing cholangitis, and autoimmune hepatitis in a Norwegian population. Scand J Gastroenterol. 1998;33(1):99-103.
6. Boonstra K, van Erpecum KJ, van Nieuwkerk KM*, et al.* Primary sclerosing cholangitis is associated with a distinct phenotype of inflammatory bowel disease. Inflamm Bowel Dis. 2012;18(12):2270-6.
7. Broome U, Olsson R, Loof L*, et al.* Natural history and prognostic factors in 305 Swedish patients with primary sclerosing cholangitis. Gut. 1996;38(4):610-5.
8. Escorsell A, Pares A, Rodes J*, et al.* Epidemiology of primary sclerosing cholangitis in Spain. Spanish Association for the Study of the Liver. J Hepatol. 1994;21(5):787-91.
9. Okolicsanyi L, Fabris L, Viaggi S*, et al.* Primary sclerosing cholangitis: clinical presentation, natural history and prognostic variables: an Italian multicentre study. The Italian PSC Study Group. Eur J Gastroenterol Hepatol. 1996;8(7):685-91.
10. Takikawa H, Takamori Y, Tanaka A*, et al.* Analysis of 388 cases of primary sclerosing cholangitis in Japan; Presence of a subgroup without pancreatic involvement in older patients. Hepatol Res. 2004;29(3):153-9.
11. Saarinen S, Olerup O, Broome U. Increased frequency of autoimmune diseases in patients with primary sclerosing cholangitis. Am J Gastroenterol. 2000;95(11):3195-9.
12. Burak K, Angulo P, Pasha TM*, et al.* Incidence and risk factors for cholangiocarcinoma in primary sclerosing cholangitis. Am J Gastroenterol. 2004;99(3):523-6.
13. Boberg KM, Bergquist A, Mitchell S*, et al.* Cholangiocarcinoma in primary sclerosing cholangitis: risk factors and clinical presentation. Scand J Gastroenterol. 2002;37(10):1205-11.
14. de Valle MB, Bjornsson E, Lindkvist B. Mortality and cancer risk related to primary sclerosing cholangitis in a Swedish population-based cohort. Liver Int. 2012;32(3):441-8.
15. Bergquist A, Ekbom A, Olsson R*, et al.* Hepatic and extrahepatic malignancies in primary sclerosing cholangitis. J Hepatol. 2002;36(3):321-7.
16. Claessen MM, Vleggaar FP, Tytgat KM*, et al.* High lifetime risk of cancer in primary sclerosing cholangitis. J Hepatol. 2009;50(1):158-64.
17. Yimam KK, Bowlus CL. Diagnosis and classification of primary sclerosing cholangitis. Autoimmun Rev. 2014;13(4-5):445-50.
18. Kaplan GG, Laupland KB, Butzner D*, et al.* The burden of large and small duct primary sclerosing cholangitis in adults and children: a population-based analysis. Am J Gastroenterol. 2007;102(5):1042-9.
19. European Association for the Study of the L. EASL Clinical Practice Guidelines: Management of cholestatic liver diseases. J Hepatol. 2009;51(2):237-67.
20. Tischendorf JJ, Hecker H, Kruger M*, et al.* Characterization, outcome, and prognosis in 273 patients with primary sclerosing cholangitis: A single center study. Am J Gastroenterol. 2007;102(1):107-14.
21. Chapman R, Fevery J, Kalloo A*, et al.* Diagnosis and management of primary sclerosing cholangitis. Hepatology. 2010;51(2):660-78.

22. Ponsioen CY, Vrouenraets SM, Prawirodirdjo W*, et al.* Natural history of primary sclerosing cholangitis and prognostic value of cholangiography in a Dutch population. Gut. 2002;51(4):562-6.
23. Graziadei IW, Wiesner RH, Marotta PJ*, et al.* Long-term results of patients undergoing liver transplantation for primary sclerosing cholangitis. Hepatology. 1999;30(5):1121-7.
24. Brandsaeter B, Friman S, Broome U*, et al.* Outcome following liver transplantation for primary sclerosing cholangitis in the Nordic countries. Scand J Gastroenterol. 2003;38(11):1176-83.
25. Fosby B, Karlsen TH, Melum E. Recurrence and rejection in liver transplantation for primary sclerosing cholangitis. World J Gastroenterol. 2012;18(1):1-15.
26. Cullen SN, Rust C, Fleming K*, et al.* High dose ursodeoxycholic acid for the treatment of primary sclerosing cholangitis is safe and effective. J Hepatol. 2008;48(5):792-800.
27. Lindor KD. Ursodiol for primary sclerosing cholangitis. Mayo Primary Sclerosing Cholangitis-Ursodeoxycholic Acid Study Group. N Engl J Med. 1997;336(10):691-5.
28. Mitchell SA, Bansi DS, Hunt N*, et al.* A preliminary trial of high-dose ursodeoxycholic acid in primary sclerosing cholangitis. Gastroenterology. 2001;121(4):900-7.
29. Olsson R, Boberg KM, de Muckadell OS*, et al.* High-dose ursodeoxycholic acid in primary sclerosing cholangitis: a 5-year multicenter, randomized, controlled study. Gastroenterology. 2005;129(5):1464-72.
30. Lindor KD, Kowdley KV, Luketic VA*, et al.* High-dose ursodeoxycholic acid for the treatment of primary sclerosing cholangitis. Hepatology. 2009;50(3):808-14.
31. Angulo P, Maor-Kendler Y, Lindor KD. Small-duct primary sclerosing cholangitis: A long-term follow-up study. Hepatology. 2002;35(6):1494-500.
32. Bjornsson E, Boberg KM, Cullen S*, et al.* Patients with small duct primary sclerosing cholangitis have a favourable long term prognosis. Gut. 2002;51(5):731-5.
33. Broome U, Glaumann H, Lindstom E*, et al.* Natural history and outcome in 32 Swedish patients with small duct primary sclerosing cholangitis (PSC). J Hepatol. 2002;36(5):586-9.
34. Bjornsson E, Olsson R, Bergquist A*, et al.* The natural history of small-duct primary sclerosing cholangitis. Gastroenterology. 2008;134(4):975-80.
35. Lai J, Taouli B, Iyer KR*, et al.* Cholangiolocellular carcinoma in a pediatric patient with small duct sclerosing cholangitis: a case report. Semin Liver Dis. 2012;32(4):360-6.
36. Mitchell SA, Thyssen M, Orchard TR*, et al.* Cigarette smoking, appendectomy, and tonsillectomy as risk factors for the development of primary sclerosing cholangitis: a case control study. Gut. 2002;51(4):567-73.
37. Loftus EV, Jr., Sandborn WJ, Tremaine WJ*, et al.* Primary sclerosing cholangitis is associated with nonsmoking: a case-control study. Gastroenterology. 1996;110(5):1496-502.
38. van Erpecum KJ, Smits SJ, van de Meeberg PC*, et al.* Risk of primary sclerosing cholangitis is associated with nonsmoking behavior. Gastroenterology. 1996;110(5):1503-6.
39. Andersen IM, Tengesdal G, Lie BA*, et al.* Effects of coffee consumption, smoking, and hormones on risk for primary sclerosing cholangitis. Clin Gastroenterol Hepatol. 2014;12(6):1019-28.
40. Karlsen TH, Franke A, Melum E*, et al.* Genome-Wide Association Analysis in Primary Sclerosing Cholangitis. Gastroenterology. 2010;138(3):1102-11.
41. Melum E, Franke A, Schramm C*, et al.* Genome-wide association analysis in primary sclerosing cholangitis identifies two non-HLA susceptibility loci. Nat Genet. 2011;43(1):17-9.
42. Janse M, Lamberts LE, Franke L*, et al.* Three ulcerative colitis susceptibility loci are associated with primary sclerosing cholangitis and indicate a role for IL2, REL, and CARD9. Hepatology. 2011;53(6):1977-85.
43. Ellinghaus D, Folseraas T, Holm K*, et al.* Genome-wide association analysis in primary sclerosing cholangitis and ulcerative colitis identifies risk loci at GPR35 and TCF4. Hepatology. 2013;58(3):1074-83.

44. Folseraas T, Melum E, Rausch P*, et al.* Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. J Hepatol. 2012;57(2):366-75.
45. Liu JZ, Hov JR, Folseraas T*, et al.* Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. Nat Genet. 2013;45(6):670-5.
46. Adams DH, Eksteen B. Aberrant homing of mucosal T cells and extra-intestinal manifestations of inflammatory bowel disease. Nat Rev Immunol. 2006;6(3):244-51.
47. O'Mahony CA, Vierling JM. Etiopathogenesis of primary sclerosing cholangitis. Semin Liver Dis. 2006;26(1):3-21.
48. Popov Y, Patsenker E, Fickert P*, et al.* Mdr2 (Abcb4)-/- mice spontaneously develop severe biliary fibrosis via massive dysregulation of pro- and antifibrogenic genes. Journal of Hepatology. 2005;43(6):1045-54.
49. Trauner M, Fickert P, Wagner M. MDR3 (ABCB4) defects: a paradigm for the genetics of adult cholestatic syndromes. Semin Liver Dis. 2007;27(1):77-98.
50. Bergquist A, Montgomery SM, Bahmanyar S*, et al.* Increased risk of primary sclerosing cholangitis and ulcerative colitis in first-degree relatives of patients with primary sclerosing cholangitis. Clin Gastroenterol Hepatol. 2008;6(8):939-43.
51. Folseraas T, Liaskou E, Anderson CA*, et al.* Genetics in PSC: what do the "risk genes" teach us? Clin Rev Allergy Immunol. 2015;48(2-3):154-64.
52. Venter JC, Adams MD, Myers EW*, et al.* The sequence of the human genome. Science. 2001;291(5507):1304-51.
53. Lander ES, Linton LM, Birren B*, et al.* Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
54. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006;7(2):85-97.
55. Weischenfeldt J, Symmons O, Spitz F*, et al.* Phenotypic impact of genomic structural variation: insights from and for human disease. Nat Rev Genet. 2013;14(2):125-38.
56. Watson JD, Crick FH. The structure of DNA. Cold Spring Harb Symp Quant Biol. 1953;18:123-31.
57. International HapMap C, Frazer KA, Ballinger DG*, et al.* A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449(7164):851-61.
58. International HapMap C. A haplotype map of the human genome. Nature. 2005;437(7063):1299-320.
59. Genomes Project C, Abecasis GR, Altshuler D*, et al.* A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73.
60. Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. Annual review of genomics and human genetics. 2013;14:301-23.
61. Horton R, Wilming L, Rand V*, et al.* Gene map of the extended human MHC. Nat Rev Genet. 2004;5(12):889-99.
62. Flajnik MF, Canel C, Kramer J*, et al.* Which came first, MHC class I or class II? Immunogenetics. 1991;33(5-6):295-300.
63. Trowsdale J. Genetic and functional relationships between MHC and NK receptor genes. Immunity. 2001;15(3):363-74.
64. Brown D, Trowsdale J, Allen R. The LILR family: modulators of innate and adaptive immune pathways in health and disease. Tissue Antigens. 2004;64(3):215-25.
65. Dausset J. [Iso-leuko-antibodies]. Acta haematologica. 1958;20(1-4):156-66.
66. Thorsby E. A short history of HLA. Tissue Antigens. 2009;74(2):101-16.
67. Ansari D, Bucin D, Nilsson J. Human leukocyte antigen matching in heart transplantation: systematic review and meta-analysis. Transpl Int. 2014;27(8):793-804.
68. Susal C, Opelz G. Current role of human leukocyte antigen matching in kidney transplantation. Current opinion in organ transplantation. 2013;18(4):438-44.
69. Lan X, Zhang MM, Pu CL*, et al.* Impact of human leukocyte antigen mismatching on outcomes of liver transplantation: a meta-analysis. World J Gastroenterol. 2010;16(27):3457-64.

70. Fisher LR, Henley KS, Lucey MR. Acute cellular rejection after liver transplantation: variability, morbidity, and mortality. Liver Transpl Surg. 1995;1(1):10-5.
71. Shaked A, Ghobrial RM, Merion RM*, et al.* Incidence and severity of acute cellular rejection in recipients undergoing adult living donor or deceased donor liver transplantation. Am J Transplant. 2009;9(2):301-8.
72. Hanvesakul R, Spencer N, Cook M*, et al.* Donor HLA-C genotype has a profound impact on the clinical outcome following liver transplantation. Am J Transplant. 2008;8(9):1931-41.
73. Tran TH, Middleton D, Dohler B*, et al.* Reassessing the impact of donor HLA-C genotype on long-term liver transplant survival. Am J Transplant. 2009;9(7):1674-8.
74. Price P, Witt C, Allcock R*, et al.* The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases. Immunol Rev. 1999;167:257-74.
75. Dilthey A, Leslie S, Moutsianas L*, et al.* Multi-population classical HLA type imputation. PLoS computational biology. 2013;9(2):e1002877.
76. Jia X, Han B, Onengut-Gumuscu S*, et al.* Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One. 2013;8(6):e64683.
77. Donaldson PT, Farrant JM, Wilkinson ML*, et al.* Dual association of HLA DR2 and DR3 with primary sclerosing cholangitis. Hepatology. 1991;13(1):129-33.
78. Donaldson PT, Norris S. Evaluation of the role of MHC class II alleles, haplotypes and selected amino acid sequences in primary sclerosing cholangitis. Autoimmunity. 2002;35(8):555-64.
79. Farrant JM, Doherty DG, Donaldson PT*, et al.* Amino acid substitutions at position 38 of the DR beta polypeptide confer susceptibility to and protection from primary sclerosing cholangitis. Hepatology. 1992;16(2):390-5.
80. Hov JR, Lleo A, Selmi C*, et al.* Genetic associations in Italian primary sclerosing cholangitis: Heterogeneity across Europe defines a critical role for HLA-C. J Hepatol. 2010;52(5):712-7.
81. Karlsen TH, Boberg KM, Vatn M*, et al.* Different HLA class II associations in ulcerative colitis patients with and without primary sclerosing cholangitis. Genes Immun. 2007;8(3):275-8.
82. Leidenius MH, Koskimies SA, Kellokumpu IH*, et al.* HLA antigens in ulcerative colitis and primary sclerosing cholangitis. APMIS : acta pathologica, microbiologica, et immunologica Scandinavica. 1995;103(7-8):519-24.
83. Moloney MM, Thomson LJ, Strettell MJ*, et al.* Human leukocyte antigen-C genes and susceptibility to primary sclerosing cholangitis. Hepatology. 1998;28(3):660-2.
84. Noguchi K, Kobayashi M, Yagihashi A*, et al.* HLA antigens in primary sclerosing cholangitis. Transplant Proc. 1992;24(6):2775-6.
85. Spurkland A, Saarinen S, Boberg KM*, et al.* HLA class II haplotypes in primary sclerosing cholangitis patients from five European populations. Tissue Antigens. 1999;53(5):459-69.
86. Wiencke K, Karlsen TH, Boberg KM*, et al.* Primary sclerosing cholangitis is associated with extended HLA-DR3 and HLA-DR6 haplotypes. Tissue Antigens. 2007;69(2):161-9.
87. Wiencke K, Spurkland A, Schrumpf E*, et al.* Primary sclerosing cholangitis is associated to an extended B8-DR3 haplotype including particular MICA and MICB alleles. Hepatology. 2001;34(4 Pt 1):625-30.
88. Howson JM, Walker NM, Clayton D*, et al.* Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. Diabetes, obesity & metabolism. 2009;11 Suppl 1:31-45.
89. Cucca F, Lampis R, Congia M*, et al.* A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes and the structure of their proteins. Hum Mol Genet. 2001;10(19):2025-37.
90. Sollid LM, Jabri B. Triggers and drivers of autoimmunity: lessons from coeliac disease. Nat Rev Immunol. 2013;13(4):294-302.
91. Invernizzi P, Ransom M, Raychaudhuri S*, et al.* Classical HLA-DRB1 and DPB1 alleles account for HLA associations with primary biliary cirrhosis. Genes Immun. 2012;13(6):461-8.

92. Invernizzi P. Human leukocyte antigen in primary biliary cirrhosis: an old story now reviving. Hepatology. 2011;54(2):714-23.
93. Goyette P, Boucher G, Mallon D*, et al.* High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nat Genet. 2015;47(2):172-9.
94. Raychaudhuri S, Sandor C, Stahl EA*, et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet. 2012;44(3):291-6.
95. Vignal C, Bansal AT, Balding DJ*, et al.* Genetic association of the major histocompatibility complex with rheumatoid arthritis implicates two non-DRB1 loci. Arthritis Rheum. 2009;60(1):53-62.
96. Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum. 1987;30(11):1205-13.
97. International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S*, et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9.
98. Diaz-Pena R, Aransay AM, Bruges-Armas J*, et al.* Fine mapping of a major histocompatibility complex in ankylosing spondylitis: association of the HLA-DPA1 and HLA-DPB1 regions. Arthritis Rheum. 2011;63(11):3305-12.
99. Gregersen PK, Kosoy R, Lee AT*, et al.* Risk for myasthenia gravis maps to a (151) Pro-->Ala change in TNIP1 and to human leukocyte antigen-B*08. Ann Neurol. 2012;72(6):927-35.
100. Maniaol AH, Elsais A, Lorentzen AR*, et al.* Late onset myasthenia gravis is associated with HLA DRB1*15:01 in the Norwegian population. PLoS One. 2012;7(5):e36603.
101. Faraco J, Lin L, Kornum BR*, et al.* ImmunoChip study implicates antigen presentation to T cells in narcolepsy. PLoS Genet. 2013;9(2):e1003270.
102. Hor H, Kutalik Z, Dauvilliers Y*, et al.* Genome-wide association study identifies new HLA class II haplotypes strongly protective against narcolepsy. Nat Genet. 2010;42(9):786-9.
103. Rahman P, Elder JT. Genetics of psoriasis and psoriatic arthritis: a report from the GRAPPA 2010 annual meeting. J Rheumatol. 2012;39(2):431-3.
104. International Genetics of Ankylosing Spondylitis C, Cortes A, Hadler J*, et al.* Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. Nat Genet. 2013;45(7):730-8.
105. Mackie SL, Taylor JC, Martin SG*, et al.* A spectrum of susceptibility to rheumatoid arthritis within HLA-DRB1: stratification by autoantibody status in a large UK population. Genes Immun. 2012;13(2):120-8.
106. Karlsen TH, Boberg KM, Olsson M*, et al.* Particular genetic variants of ligands for natural killer cell receptors may contribute to the HLA associated risk of primary sclerosing cholangitis. J Hepatol. 2007;46(5):899-906.
107. Norris S, Kondeatis E, Collins R*, et al.* Mapping MHC-encoded susceptibility and resistance in primary sclerosing cholangitis: the role of MICA polymorphism. Gastroenterology. 2001;120(6):1475-82.
108. Mitchell SA, Grove J, Spurkland A*, et al.* Association of the tumour necrosis factor alpha -308 but not the interleukin 10 -627 promoter polymorphism with genetic susceptibility to primary sclerosing cholangitis. Gut. 2001;49(2):288-94.
109. Bernal W, Moloney M, Underhill J*, et al.* Association of tumor necrosis factor polymorphism with primary sclerosing cholangitis. J Hepatol. 1999;30(2):237-41.
110. Bowlus CL, Li C-S, Karlsen TH*, et al.* Primary sclerosing cholangitis in genetically diverse populations listed for liver transplantation: Unique clinical and human leukocyte antigen associations. Liver Transplantation. 2010;16(11):1324-30.
111. Hov JER. Functional genetics in primary sclerosing cholangitis: Studies of the bile acid receptor TGR5 and genes in the HLA complex. Oslo: Unipub; 2011.

112. Hov JR, Kosmoliaptsis V, Traherne JA*, et al.* Electrostatic modifications of the human leukocyte antigen-DR P9 peptide-binding pocket and susceptibility to primary sclerosing cholangitis. Hepatology. 2011;53(6):1967-76.
113. Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. Lancet. 2001;358(9290):1356-60.
114. Bjornsson E. Small-duct primary sclerosing cholangitis. Curr Gastroenterol Rep. 2009;11(1):37-41.
115. Banff schema for grading liver allograft rejection: an international consensus document. Hepatology. 1997;25(3):658-63.
116. Lennard-Jones JE. Classification of inflammatory bowel disease. Scand J Gastroenterol Suppl. 1989;170:2-6; discussion 16-9.
117. McCarthy MI, Abecasis GR, Cardon LR*, et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9(5):356-69.
118. Zondervan KT, Cardon LR. Designing candidate gene and genome-wide case-control association studies. Nature protocols. 2007;2(10):2492-501.
119. Thomson G. Mapping disease genes: family-based association studies. Am J Hum Genet. 1995;57(2):487-98.
120. Garner C. The use of random controls in genetic association studies. Human heredity. 2006;61(1):22-6.
121. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. Lancet. 2003;361(9357):598-604.
122. Visvikis S, Schlenck A, Maurice M. DNA extraction and stability for epidemiological studies. Clinical chemistry and laboratory medicine : CCLM / FESCC. 1998;36(8):551-5.
123. Lovmar L, Syvanen AC. Multiple displacement amplification to create a long-lasting source of DNA for genetic studies. Hum Mutat. 2006;27(7):603-14.
124. Adams SD, Barracchini KC, Chen D*, et al.* Ambiguous allele combinations in HLA Class I and Class II sequence-based typing: when precise nucleotide sequencing leads to imprecise allele identification. Journal of translational medicine. 2004;2(1):30.
125. Parham P. MHC class I molecules and KIRs in human history, health and survival. Nat Rev Immunol. 2005;5(3):201-14.
126. Muller CA, Engler-Blum G, Gekeler V*, et al.* Genetic and serological heterogeneity of the supertypic HLA-B locus specificities Bw4 and Bw6. Immunogenetics. 1989;30(3):200-7.
127. Martin MP, Gao X, Lee JH*, et al.* Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. Nat Genet. 2002;31(4):429-34.
128. Tobler AR, Short S, Andersen MR*, et al.* The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. Journal of biomolecular techniques : JBT. 2005;16(4):398-406.
129. De la Vega FM, Lazaruk KD, Rhodes MD*, et al.* Assessment of two flexible and compatible SNP genotyping platforms: TaqMan SNP Genotyping Assays and the SNPlex Genotyping System. Mutat Res. 2005;573(1-2):111-35.
130. Kraft P, Zeggini E, Ioannidis JP. Replication in genome-wide association studies. Statistical science : a review journal of the Institute of Mathematical Statistics. 2009;24(4):561-73.
131. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nature protocols. 2008;3(6):1101-8.
132. Jiang W, Johnson C, Jayaraman J*, et al.* Copy number variation leads to considerable diversity for B but not A haplotypes of the human KIR genes encoding NK cell receptors. Genome Res. 2012;22(10):1845-54.
133. Anderson CA, Pettersson FH, Clarke GM*, et al.* Data quality control in genetic case-control association studies. Nature protocols. 2010;5(9):1564-73.
134. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet. 2005;76(5):887-93.
135. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010;11(7):499-511.

136. Li Y, Willer CJ, Ding J*, et al.* MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genetic epidemiology. 2010;34(8):816-34.
137. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. Bioinformatics. 2015;31(5):782-4.
138. Howie B, Fuchsberger C, Stephens M*, et al.* Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44(8):955-9.
139. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3. 2011;1(6):457-70.
140. Tiwari J, Terasaki P. The Data and Statistical Analysis. HLA and Disease Associations: Springer New York; 1985. p. 18-27.
141. Clarke GM, Anderson CA, Pettersson FH*, et al.* Basic statistical analysis in genetic case-control studies. Nature protocols. 2011;6(2):121-33.
142. Cordell HJ, Clayton DG. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. Am J Hum Genet. 2002;70(1):124-41.
143. Dupont WD, Plummer WD, Jr. Power and sample size calculations. A review and computer program. Controlled clinical trials. 1990;11(2):116-28.
144. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 2001;68(4):978-89.
145. Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet. 2005;76(3):449-62.
146. So HC, Gui AH, Cherny SS*, et al.* Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genetic epidemiology. 2011;35(5):310-7.
147. Noble JA, Valdes AM. Genetics of the HLA region in the prediction of type 1 diabetes. Current diabetes reports. 2011;11(6):533-42.
148. Molberg O, Kett K, Scott H*, et al.* Gliadin specific, HLA DQ2-restricted T cells are commonly found in small intestinal biopsies from coeliac disease patients, but not from controls. Scand J Immunol. 1997;46(3):103-9.
149. Khakoo SI, Carrington M. KIR and disease: a model system or system of models? Immunol Rev. 2006;214:186-201.
150. Berglin L, Bergquist A, Johansson H*, et al.* In situ characterization of intrahepatic non-parenchymal cells in PSC reveals phenotypic patterns associated with disease severity. PLoS One. 2014;9(8):e105375.
151. Whiteside TL, Lasky S, Si L*, et al.* Immunologic analysis of mononuclear cells in liver tissues and blood of patients with primary sclerosing cholangitis. Hepatology. 1985;5(3):468-74.
152. Naess S, Lie BA, Melum E*, et al.* Refinement of the MHC risk map in a scandinavian primary sclerosing cholangitis population. PLoS One. 2014;9(12):e114486.
153. Ramos M, Lopez de Castro JA. HLA-B27 and the pathogenesis of spondyloarthritis. Tissue Antigens. 2002;60(3):191-205.
154. Gregersen JW, Holmes S, Fugger L. Humanized animal models for autoimmune diseases. Tissue Antigens. 2004;63(5):383-94.
155. Hammer RE, Maika SD, Richardson JA*, et al.* Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human beta 2m: an animal model of HLA-B27-associated human disorders. Cell. 1990;63(5):1099-112.
156. Wen L, Wong FS, Tang J*, et al.* In vivo evidence for the contribution of human histocompatibility leukocyte antigen (HLA)-DQ molecules to the development of diabetes. J Exp Med. 2000;191(1):97-104.
157. Zachary AA, Bias WB, Johnson A*, et al.* Antigen, allele, and haplotype frequencies report of the ASHI minority antigens workshops: part 1, African-Americans. Human immunology. 2001;62(10):1127-36.
158. Shayevitz C, Cohen OS, Faraone SV*, et al.* A re-review of the association between the NOTCH4 locus and schizophrenia. American journal of medical genetics Part B,

Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2012;159B(5):477-83.
159. Valdes AM, Thomson G, Type 1 Diabetes Genetics C. Several loci in the HLA class III region are associated with T1D risk after adjusting for DRB1-DQB1. Diabetes, obesity & metabolism. 2009;11 Suppl 1:46-52.
160. Horton R, Gibson R, Coggill P*, et al.* Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. Immunogenetics. 2008;60(1):1-18.
161. Candore G, Lio D, Colonna Romano G*, et al.* Pathogenesis of autoimmune diseases associated with 8.1 ancestral haplotype: effect of multiple gene interactions. Autoimmun Rev. 2002;1(1-2):29-35.
162. Benito de Valle M, Muller T, Bjornsson E*, et al.* The impact of elevated serum IgG4 levels in patients with primary sclerosing cholangitis. Dig Liver Dis. 2014;46(10):903-8.
163. Berntsen NL, Klingenberg O, Juran BD*, et al.* Association Between HLA Haplotypes and Increased Serum Levels of IgG4 in Patients With Primary Sclerosing Cholangitis. Gastroenterology. 2015;148(5):924-7 e2.
164. Hov JR, Boberg KM, Karlsen TH. Autoantibodies in primary sclerosing cholangitis. World journal of gastroenterology : WJG. 2008;14(24):3781-91.
165. Mehal WZ, Lo YM, Wordsworth BP*, et al.* HLA DR4 is a marker for rapid disease progression in primary sclerosing cholangitis. Gastroenterology. 1994;106(1):160-7.
166. Olerup O, Olsson R, Hultcrantz R*, et al.* HLA-DR and HLA-DQ are not markers for rapid disease progression in primary sclerosing cholangitis. Gastroenterology. 1995;108(3):870-8.
167. Boberg KM, Spurkland A, Rocca G*, et al.* The HLA-DR3,DQ2 heterozygous genotype is associated with an accelerated progression of primary sclerosing cholangitis. Scand J Gastroenterol. 2001;36(8):886-90.
168. van der Woude D, Lie BA, Lundstrom E*, et al.* Protection against anti-citrullinated protein antibody-positive rheumatoid arthritis is predominantly associated with HLA-DRB1*1301: a meta-analysis of HLA-DRB1 associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in four European populations. Arthritis Rheum. 2010;62(5):1236-45.
169. Terminology for hepatic allograft rejection. International Working Party. Hepatology. 1995;22(2):648-54.
170. Knechtle SJ, Kwun J. Unique aspects of rejection and tolerance in liver transplantation. Semin Liver Dis. 2009;29(1):91-101.
171. Farges O, Saliba F, Farhamant H*, et al.* Incidence of rejection and infection after liver transplantation as a function of the primary disease: possible influence of alcohol and polyclonal immunoglobulins. Hepatology. 1996;23(2):240-8.
172. Wiesner RH, Demetris AJ, Belle SH*, et al.* Acute hepatic allograft rejection: incidence, risk factors, and impact on outcome. Hepatology. 1998;28(3):638-45.
173. Oliveira LC, Porta G, Marin ML*, et al.* Autoimmune hepatitis, HLA and extended haplotypes. Autoimmun Rev. 2011;10(4):189-93.
174. Doran TJ, Derley L, Chapman J*, et al.* Severity of liver transplantation rejection is associated with recipient HLA type. Transplant Proc. 1992;24(1):192-3.
175. Doran TJ, Geczy AF, Painter D*, et al.* A large, single center investigation of the immunogenetic factors affecting liver transplantation. Transplantation. 2000;69(7):1491-8.
176. Neumann UP, Guckelberger O, Langrehr JM*, et al.* Impact of human leukocyte antigen matching in liver transplantation. Transplantation. 2003;75(1):132-7.
177. Suehiro T, Shimada M, Kishikawa K*, et al.* Influence of HLA compatibility and lymphocyte cross-matching on acute cellular rejection following living donor adult liver transplantation. Liver Int. 2005;25(6):1182-8.
178. Nikaein A, Backman L, Jennings L*, et al.* HLA compatibility and liver transplant outcome. Improved patient survival by HLA and cross-matching. Transplantation. 1994;58(7):786-92.

179. Donaldson P, Underhill J, Doherty D*, et al.* Influence of human leukocyte antigen matching on liver allograft survival and rejection: "the dualistic effect". Hepatology. 1993;17(6):1008-15.
180. Markus BH, Duquesnoy RJ, Gordon RD*, et al.* Histocompatibility and liver transplant outcome. Does HLA exert a dualistic effect? Transplantation. 1988;46(3):372-7.
181. Opelz G, Wujciak T, Dohler B*, et al.* HLA compatibility and organ transplant survival. Collaborative Transplant Study. Reviews in immunogenetics. 1999;1(3):334-42.
182. Wittig M, Anmarkrud JA, Kassens JC*, et al.* Development of a high-resolution NGS-based HLA-typing and analysis pipeline. Nucleic Acids Res. 2015;43(11):e70.

8 ERRATA

Paper I:

Page 11, Figure 3, figure box

The paper reads: "DQB1*04-DRB1*03:02"

The correct should be: "**DRB1***04-DQB1*03:02"

Paper II:

Page 7, left column, line 17

The paper reads: "DRB1*13*01"

The correct should be: "DRB1***13:01**"

Paper III:

Corrections in paper text (corrections are marked in bold):

Page 3986, right column, line 25 and 26

The paper reads: "40 % of them had AR *vs.* 24 % in the non-PSC group ($P = 0.04$)."

The correct should be: "**41%** of them had AR *vs.* **23 %** in the non-PSC group ($P =$ **0.03**)."

Page 3986, right column, line 30- 32

The paper reads: "HLA-C*07 ($P = 0.001$, OR = 2.4; 95 % CI: 1.4-4.0) and HLA-DRB1*03 ($P = 0.03$, OR = 1.9; 95 % CI: 1.0-3.3)"

The correct should be: "HLA-C*07 ($P = 0.001$, OR = **2.5**; 95 % CI: 1.4-**4.1**) and HLA-DRB1*03 ($P =$ **0.04**, OR = 1.9; 95 % CI: 1.0-3.3)"

Page 3987, left column, line 1-5

The paper reads: "HLA-DRB1*04 (P = 0.001, OR = 0.2; 95 % CI: 0.1-0.5). For HLA-B*08, HLA-C*07 and DRB1*04 the associations remained evident in a subgroup analysis of non-PSC recipients (P = 0.04, P = 0.003 and P = 0.02, respectively)."

The correct should be: "HLA-DRB1*04 (P = **0.0005**, OR = 0.2; 95 % CI: 0.1-0.5). For HLA-B*08, HLA-C*07 and DRB1*04 the associations remained evident in a subgroup analysis of non-PSC recipients (P = 0.04, P = 0.003 and P = **0.01**, respectively)."

Page 3988, left column, line 24

The paper reads: "liver recipients (n = 45)"

The correct should be: "liver recipients **(n = 44)**"

Page 3989, right column, line 1

The paper reads: "from the KIR association analysis"

The correct should be: "from the KIR **copy number variation** association analysis"

Page 3989, right column, line 5-8

The paper reads: "study population was 28 %, with a significantly higher frequency noted in patients with PSC as compared with patients transplanted on the basis of other liver diseases (40 % *vs.* 24 %, OR = 2.2, 95 % CI: 1.0-4.6, P = 0.04)"

The correct should be: "study population was **29 %**, with a significantly higher frequency noted in patients with PSC as compared with patients transplanted on the basis of other liver diseases **(41% *vs* 23 %, OR = 2.3, 95 % CI: 1.1-4.9, *P* = 0.03)**"

Page 3989, right column, line 27-32

The paper reads: "HLA-C*07 (*P* = 0.001, OR = 2.4, 95 % CI: 1.4-4.0) and HLA-DRB1*03 (*P* = 0.03, OR = 1.9, 95 % CI: 1.0-3.3) (Tables 2-4). For the HLA-DRB1*04 allele, previously shown to associate with decreased risk of PSC, reduced risk of AR in the overall population was found (*P* = 0.001, OR = 0.2, 95 % CI: 0.1-0.5) (Table 4)."

The correct should be: "HLA-C*07 (*P* = 0.001, OR = **2.5**, 95 % CI: 1.4-**4.1**) and HLA-DRB1*03 (*P* = **0.04**, OR = 1.9, 95 % CI: 1.0-3.3) (Tables 2-4). For the HLA-DRB1*04 allele, previously shown to associate with decreased risk of PSC, reduced risk of AR in the overall population was found (*P* = **0.0005**, OR = 0.2, 95 % CI: 0.1-0.5) (Table 4)."

Page 3989, right column, line 36

The paper reads: "and *P* = 0.02, respectively"

The correct should be: "and *P* = **0.01**, respectively"

Page 3992, left column, line 1

The paper reads: "(Tables 2 and 10)"

The correct should be: "(Tables **10** and **11**)"

Page 3995, right column, line 1

The paper reads: "AR frequency of 28 %"

The correct should be: "AR frequency of **29 %**"

Corrections of paper tables:

Corrections have been made in Table 2-11. New tables are presented, except for Table 10, in which the only change is change of *P*-value for *2DS4DEL* from 0.03 to 0.04. Corrections of numbers (*n*) in the various groups are marked in bold in the respective table titles. The correct number of acute rejection and non-acute rejection in the total population (*n* =143) is *n* = 41 and *n* = 102, respectively. The correct number of acute rejection and non-acute rejection in the non-PSC population (*n* = 99) is *n* = 23 and *n* = 76, respectively. The correct number of acute rejection and non-acute rejection in the PSC population (*n* = 44) is *n* = 18 and *n* = 26, respectively. Corrections of number of HLA alleles in the various groups in Table 2-7 are marked in bold. Due to these changes, allele frequencies, ORs with 95 % CI and *P*-values have been recalculated for all alleles at all loci, new results are presented in Table 2-7. In Table 8, new *P*-values are presented for *HLA-B* and *HLA-DR* (marked in bold), according to Fisher's Exact test, as this was not done in the original table. Correction of number of individuals in the PSC and non-PSC group with acute rejection has been made in Table 9, changes are marked in bold. % and *P*-values have been recalculated accordingly. In Table 11 corrections have been made for the following KIR genes; *2DL1, 2DS3, 2DS4DEL, 3DL2* and *3DS1*, changes are marked in bold.

Table 2. Frequencies of human leukocyte antigen-B alleles in the acute cellular rejection group (n = 40) **(missing n = 1)** and the non-acute cellular rejection group (n = 102) in the total population.

	Acute rejection		No acute rejection				
***HLA-B* allele**	***n* (alleles)**	**(%)**	***n* (alleles)**	**(%)**	**OR**	**95 % CI**	**Uncorrected *P* value**
*05	0	(0)	8	(3.9)	0.14	0.02-1.15	0.111
*07	10	(12.5)	23	(11.3)	1.15	0.54-2.47	0.772
*08	27	(33.8)	34	(16.7)	2.54	1.42-4.55	0.002
*12	12	(15.0)	27	(13.2)	1.18	0.58-2.41	0.698
*13	1	(1.3)	5	(2.5)	0.68	0.14-3.46	1
*14	2	(2.5)	4	(2.0)	1.42	0.33-6.08	0.676
*15	**5**	(6.3)	28	(13.7)	0.45	0.18-1.13	0.077
*16	1	(1.3)	4	(2.0)	0.84	0.16-4.42	1
*17	1	(1.3)	6	(2.9)	0.58	0.12-2.83	0.677
*18	1	(1.3)	6	(2.9)	0.58	0.12-2.83	0.677
*21	0	(0)	2	(1.0)	0.50	0.05-4.91	1
*22	3	(3.8)	2	(1.0)	3.66	0.80-16.72	0.138
*27	4	(5.0)	10	(4.9)	1.09	0.37-3.24	1
*35	4	(5.0)	19	(9.3)	0.56	0.20-1.54	0.231
*37	4	(5.0)	4	(2.0)	2.62	0.74-9.31	0.227
*40	4	(5.0)	19	(9.3)	0.56	0.20-1.54	0.231
*41	1	(1.3)	2	(1)	1.53	0.25-9.32	1
*47	0	(0)	1	(0.5)	0.84	0.08-9.42	1

OR and corresponding 95% CIs were calculated using Woolf's formula with Haldane's correction; uncorrected P-value: calculated by the χ^2 test or the Fisher's exact test where appropriate. HLA: Human leukocyte antigen; AR: acute cellular rejection; OR: odds ratio; CI: 95% confidence interval.

Table 3. Frequencies of human leukocyte antigen-C alleles in the acute cellular rejection group (n = 40) **(missing n = 1)** and the non-acute cellular rejection group (n = 102) in the total population.

	Acute rejection		**No acute rejection**				
***HLA-C* allele**	***n* (alleles)**	**(%)**	***n* (alleles)**	**(%)**	**OR**	**95 % CI**	**Uncorrected *P* value**
*01	2	(2.5)	8	(3.9)	0.74	0.19-2.79	0.730
*02	2	(2.5)	11	(5.4)	0.54	0.15-1.95	0.364
*03	10	(12.5)	43	(21.1)	0.55	0.27-1.13	0.095
*04	8	(10.0)	24	(11.8)	0.86	0.38-1.94	0.672
*05	**7**	(8.8)	18	(8.8)	1.03	0.43-2.45	0.984
*06	6	(7.5)	16	(7.8)	1.00	0.40-2.50	0.922
*07	41	(51.3)	**61**	(29.9)	2.45	1.45-4.15	0.001
*08	2	(2.5)	4	(2.0)	1.42	0.33-6.08	0.676
*12	0	(0)	8	(3.9)	0.14	0.02-1.15	0.111
*14	0	(0)	1	(0.5)	0.84	0.08-9.42	1
*15	0	(0)	4	(2.0)	0.28	0.03-2.41	0.580
*16	0	(0)	3	(1.5)	0.36	0.04-3.25	0.561
*17	**2**	(2.5)	**3**	(1.5)	1.83	0.40-8.38	0.623

OR and corresponding 95% CIs were calculated using Woolf's formula with Haldane's correction; uncorrected P-value: calculated by the χ^2 test or the Fisher's exact test where appropriate. HLA: Human leukocyte antigen; AR: acute cellular rejection; OR: odds ratio; CI: 95% confidence interval.

Table 4. Frequencies of human leukocyte antigen-DRB1 alleles in the acute cellular rejection group (n = 40) **(missing n = 1)** and the non-acute cellular rejection group **(n = 102)** in the total population.

	Acute rejection		No acute rejection				
***HLA-DRB1* allele**	***n* (alleles)**	**(%)**	***n* (alleles)**	**(%)**	**OR**	**95 % CI**	**Uncorrected *P* value**
*01	**11**	(13.8)	**22**	(10.8)	1.34	0.63-2.84	0.483
*02	15	(18.8)	31	(15.2)	1.30	0.67-2.53	0.465
*03	25	(31.3)	40	(19.6)	1.87	1.05-3.32	0.036
*04	4	(5.0)	46	(22.5)	0.20	0.08-0.52	0.0005
*07	7	(8.8)	16	(7.8)	1.17	0.48-2.82	0.801
*08	2	(2.5)	**11**	(5.4)	0.54	0.15-1.95	0.364
*09	0	(0)	1	(0.5)	0.84	0.08-9.42	1
*10	1	(1.3)	4	(2.0)	0.84	0.16-4.42	1
*11	1	(1.3)	6	(2.9)	0.58	0.12-2.83	0.677
*12	1	(1.3)	3	(1.5)	1.09	0.20-6.05	1
*13	**12**	(15.0)	21	(10.3)	1.56	0.74-3.26	0.266
*14	1	(1.3)	3	(1.5)	1.09	0.20-6.05	1

OR and corresponding 95% CIs were calculated using Woolf's formula with Haldane's correction; uncorrected P-value: calculated by the χ^2 test or the Fisher's exact test where appropriate. HLA: Human leukocyte antigen; AR: acute cellular rejection; OR: odds ratio; CI: 95% confidence interval.

Table 5. Comparison of the frequencies of human leukocyte antigen-B alleles in the acute cellular rejection group (n = 23) and the non-acute cellular rejection group (n = 76) in the non-PSC population.

	Acute rejection		Non acute rejection				
***HLA-B* allele**	***n* (alleles)**	**(%)**	***n* (alleles)**	**(%)**	**OR**	**95 % CI**	**Uncorrected *P* value**
*05	0	(0)	4	(2.6)	0.35	0.04-3.11	0.575
*07	6	(13.0)	15	(9.9)	1.42	0.55-3.70	0.586
*08	12	(26.1)	20	(13.2)	2.34	1.07-5.14	0.037
*12	**8**	(17.4)	20	(13.2)	1.43	0.60-3.37	0.470
*13	1	(2.2)	4	(2.6)	1.09	0.20-5.80	1
*14	2	(4.3)	4	(2.6)	1.85	0.43-8.06	0.625
*15	**4**	(8.7)	**22**	(14.5)	0.61	0.22-1.71	0.309
*16	1	(2.2)	3	(2.0)	1.41	0.25-7.94	1
*17	0	(0)	4	(2.6)	0.35	0.04-3.11	0.575
*18	1	(2.2)	6	(3.9)	0.74	0.15-3.70	1
*21	0	(0)	2	(1.3)	0.65	0.07-6.37	1
*22	0	(0)	2	(1.3)	0.65	0.07-6.37	1
*27	2	(4.3)	7	(4.6)	1.09	0.28-4.28	1
*35	4	(8.7)	15	(9.9)	0.94	0.33-2.71	1
*37	2	(4.3)	**4**	(2.6)	1.85	0.43-8.06	0.625
*40	2	(4.3)	17	(11.2)	0.43	0.12-1.55	0.253
*41	1	(2.2)	2	(1.3)	1.98	0.32-12.24	0.550
*47	0	(0)	1	(0.7)	1.09	0.10-12.25	1

OR and corresponding 95% CIs were calculated using Woolf's formula with Haldane's correction; uncorrected P-value: calculated by the χ^2 test or the Fisher's exact test where appropriate. HLA: Human leukocyte antigen; AR: acute cellular rejection; OR: odds ratio; CI: 95% confidence interval.

Table 6. Comparison of the frequencies of human leukocyte antigen-C alleles in the acute cellular rejection group (n = 23) and the non-acute cellular rejection group (n = 76) in the non-PSC population.

	Acute rejection		No acute rejection				
***HLA-C* allele**	***n* (alleles)**	**(%)**	***n* (alleles)**	**(%)**	**OR**	**95 % CI**	**Uncorrected *P*-value**
*01	**0**	(0)	5	(3.3)	0.29	0.03-2.46	0.592
*02	1	(2.2)	7	(4.6)	0.64	0.13-3.12	0.684
*03	4	(8.7)	34	(22.4)	0.36	0.13-0.99	0.039
*04	7	(15.2)	21	(13.8)	1.16	0.48-2.81	0.811
*05	**5**	(10.9)	12	(7.9)	1.49	0.53-4.16	0.551
*06	3	(6.5)	12	(7.9)	0.90	0.28-2.92	1
*07	23	(50.0)	40	(26.3)	2.78	1.42-5.42	0.003
*08	2	(4.3)	4	(2.6)	1.85	0.43-8.06	0.625
*12	0	(0)	6	(3.9)	0.24	0.03-2.02	0.339
*15	0	(0)	4	(2.6)	0.35	0.04-3.11	0.575
*16	0	(0)	3	(2.0)	0.46	0.05-4.21	1
*17	1	(2.2)	4	(2.6)	1.09	0.20-5.80	1

OR and corresponding 95% CIs were calculated using Woolf's formula with Haldane's correction; uncorrected P-value: calculated by the χ^2 test or the Fisher's exact test where appropriate. HLA: Human leukocyte antigen; AR: acute cellular rejection; OR: odds ratio; CI: 95% confidence interval.

Table 7. Comparison of the frequencies of human leukocyte antigen-DRB1 alleles in the acute cellular rejection group (n = 23) and the non-acute cellular rejection group (n = 76) in the non-PSC population.

	Acute rejection		No acute rejection				
***HLA-DRB1* allele**	***n* (alleles)**	**(%)**	***n* (alleles)**	**(%)**	**OR**	**95 % CI**	**Uncorrected *P*-value**
*01	7	(15.2)	18	(11.8)	1.38	0.56-3.39	0.546
*02	**7**	(15.2)	23	(15.1)	1.05	0.44-2.51	0.989
*03	13	(28.3)	24	(15.8)	2.11	0.99-4.50	0.057
*04	3	(6.5)	35	(23.0)	0.27	0.09-0.79	0.013
*07	6	(13.0)	12	(7.9)	1.80	0.68-4.82	0.378
*08	2	(4.3)	10	(6.6)	0.76	0.20-2.85	0.736
*10	0	(0)	3	(2.0)	0.46	0.05-4.21	1
*11	1	(2.2)	6	(3.9)	0.74	0.15-3.70	1
*12	1	(2.2)	3	(2.0)	1.41	0.25-7.94	1
*13	**6**	(13.0)	15	(9.9)	1.42	0.55-3.70	0.586
*14	0	(0)	3	(2.0)	0.46	0.05-4.21	1

OR and corresponding 95% CIs were calculated using Woolf's formula with Haldane's correction; uncorrected P-value: calculated by the χ^2 test or the Fisher's exact test where appropriate. HLA: Human leukocyte antigen; AR: acute cellular rejection; OR: odds ratio; CI: 95% confidence interval.

Table 8. Data on human leukocyte antigen mismatches on the risk of acute cellular rejection after liver transplantation *n* (%)

Locus	MM (*n*)	Non-AR group (*n* = 102)	AR group (*n* = 41)	*P* value
HLA-A	1-2	86 (84.3)	38 (92.7)	0.18
	0	16 (15.7)	3 (7.3)	
HLA-B	1-2	97 (95.1)	41 (100)	**0.32**
	0	5 (4.9)	0 (0)	
HLA-DR	1-2	94 (92.2)	40 (97.6)	**0.45**
	0	8 (7.8)	1 (2.4)	
A, B, DR	5-6	40 (39.2)	22 (53.7)	0.12
	0-4	62 (60.8)	19 (46.3)	

Uncorrected P values were calculated by the χ^2 test or the Fisher's exact test where appropriate. HLA: Human leukocyte antigen; MM: Mismatches; AR: Acute cellular rejection.

Table 9. Data on human leukocyte antigen mismatches on the risk of acute cellular rejection after liver transplantation, according to primary sclerosing cholangitis or non-primary sclerosing cholangitis *n* (%)

Locus	MM (*n*)	non-AR group (PSC, *n* = 26) Non-PSC, *n* = 76)	AR group (PSC, *n* = **18**) Non-PSC, *n* = **23**)	*P* value
HLA-A				
(PSC)	1-2	20 (55.6)	**16** (44.4)	0.44
	0	6 (75.0)	2 (25.0)	
(non-PSC)	1-2	65 (74.7)	**22** (25.3)	0.29
	0	11 (91.7)	1 (8.3)	
HLA-B				
(PSC)	1-2	25 (58.1)	**18** (41.9)	1
	0	1 (100)	0 (0)	
(non-PSC)	1-2	73 (76.0)	**23** (24.0)	1
	0	3 (100)	0 (0)	
HLA-DR				
(PSC)	1-2	26 (60.5)	**17** (39.5)	0.41
	0	0 (0)	1 (100)	
(non-PSC)	1-2	67 (75.3)	**22** (24.7)	0.45
	0	9 (90.0)	1 (10.0)	
A, B, DR				
(PSC)	5-6	12 (52.2)	**11** (47.8)	0.33
	0-4	14 (66.7)	7 (33.3)	
(non-PSC)	5-6	29 (69.0)	**13** (31.0)	0.12
	0-4	47 (82.5)	10 (17.5)	

Uncorrected *P*-values were calculated by the x^2 test or the Fisher's exact test where appropriate. HLA: Human leukocyte antigen; MM: mismatches; AR: acute cellular rejection.

Table 11. Data on the relationship between killer immunoglobulin-like receptor gene phenotype in the recipient and the risk of acute cellular rejection after liver transplantation in patients with primary sclerosing cholangitis compared with patients without primary sclerosing cholangitis

Recipient *KIR* gene phenotype	Incidence of AR	*P* value
2DL1		
(PSC)		-
Negative (*n* = 0)	-	
Positive (*n* = **44**)	40 %	
(other)		1.00
Negative (*n* = 3)	0 %	
Positive (*n* = 96)	25 %	
2DL2		
(PSC)		0.37
Negative (*n* = 23)	33 %	
Positive (*n* = 21)	48 %	
(other)		0.71
Negative (*n* = 54)	26 %	
Positive (*n* = 45)	22 %	
2DL3		
(PSC)		-
Negative (*n* = 0)	-	
Positive (*n* = 44)	40 %	
(other)		1.00
Negative (*n* = 10)	20 %	
Positive (*n* = 89)	24 %	
2DL4		
(PSC)		-
Negative (*n* = 0)	-	
Positive (*n* = 44)	40 %	
(other)		-
Negative (*n* = 0)	-	
Positive (*n* = 99)	24 %	
2DL5		
(PSC)		0.17
Negative (*n* = 28)	48 %	
Positive (*n* = 16)	25 %	
(other)		0.81
Negative (*n* = 51)	25 %	
Positive (*n* = 48)	22 %	
2DP1		
(PSC)		-
Negative (*n* = 0)	-	
Positive (*n* = 44)	40 %	

(other)		1.00
Negative (*n* = 3)	0 %	
Positive (*n* = 96)	25 %	
2DS1		
(PSC)		0.14
Negative (*n* = 13)	26 %	
Positive (*n* = 31)	20 %	
(other)		0.51
Negative (*n* = 60)	26 %	
Positive (*n* = 39)	20 %	
2DS2		
(PSC)		0.37
Negative (*n* = 23)	33 %	
Positive (*n* = 21)	48 %	
(other)		0.48
Negative (*n* = 54)	26 %	
Positive (*n* = 45)	20 %	
2DS3		
(PSC)		**0.59**
Negative (*n* = **33**)	**45 %**	
Positive (*n* = 11)	36 %	
(other)		**0.52**
Negative (*n* = **75**)	22 %	
Positive (*n* = **24**)	28 %	
2DS4DEL[1]		
(PSC)		0.11
Negative (*n* = 8)	38 %	
Positive (*n* = 36)	43 %	
(other)		**0.09**
Negative (*n* = 15)	7 %	
Positive (*n* = **84**)	27 %	
2DS4WT[2]		
(PSC)		0.81
Negative (*n* = 21)	38 %	
Positive (*n* = 23)	42 %	
(other)		0.71
Negative (*n* = 56)	22 %	
Positive (*n* = 43)	26 %	
2DS4TOTA[3]		
(PSC)		
Negative (*n* = 0)	-	-
Positive (*n* = 44)	40 %	
(other)		1.00

Negative (*n* = 3)	0 %	
Positive (*n* = 96)	25 %	
2DS5		
(PSC)		0.46
Negative (*n* = 34)	43 %	
Positive (*n* =10)	30 %	
(other)		0.56
Negative (*n* = 69)	25 %	
Positive (*n* = 30)	20 %	
3DL1E4[4]		
(PSC)		-
Negative (*n* = 0)	-	
Positive (*n* = 44)	40 %	
(other)		-
Negative (*n* = 0)	-	
Positive (*n* = 99)	25 %	
3DL1E9[5]		
(PSC)		-
Negative (*n* = 0)	-	
Positive (*n* = 44)	40 %	
(other)		1.00
Negative (*n* = 3)	0 %	
Positive (*n* = 96)	25 %	
3DL2		
(PSC)		-
Negative (*n* = 0)	-	
Positive (*n* = 44)	40 %	
(other)		-
Negative (*n* = 0)	-	
Positive (*n* = **99**)	24 %	
3DL3		
(PSC)		-
Negative (*n* = 0)	-	
Positive (*n* = 44)	40 %	
(other)		-
Negative (*n* = 0)	-	
Positive (*n* = 99)	24 %	
3DP1		
(PSC)		-
Negative (*n* = 0)	-	
Positive (*n* = 44)	40 %	
(other)		-
Negative (*n* = 0)	-	

Positive (*n* = 99)	24 %	
3DS1		
(PSC)		0.42
Negative (*n* = 31)	44 %	
Positive (*n* = 13)	31 %	
(other)		**0.98**
Negative (*n* = **60**)	**23 %**	
Positive (*n* =**39**)	23 %	

KIR: Killer immunoglobulin-like receptor; AR: Acute cellular rejection; *n*: No. of patients; LTX: Liver transplantation; PSC: Primary sclerosing cholangitis; [1]*2DS4DEL* refers to the 22-bp deletion variant of *2DS4*; [2]*2DS4WT* refers to the full-length form of the gene; [3]*2DS4TOTA* refers to the total number of *2DS4*, i.e. *2DS4DEL* and *2DS4WT* combined; [4,5]*3DL1E4* and *3DL1E9* refers to exon 4 and 9 of the *3DL1* gene respectively. Genes with S in the name (e.g. *KIR2DS4*) encode activating KIRs, genes with L in the name (e.g. *KIR3DL1*) encode inhibiting KIRs and genes with a P in the name (e.g *KIR3DP1*) are pseudogenes. Uncorrected *P*-value was calculated by the χ^2 test or the Fisher's exact test where appropriate.